

From: Rao, Manjunath N.
Sent: Thursday, November 07, 2002 3:37 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/917,378

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10D 01
Phone: 306-5681

Date: 11-7-02

Please search the following as soon as possible for application with serial number **09/917,378**

SEQ ID NO: 2, against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

SEQ ID NO:1, 3, 4, and 5, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Crystal Mall 1, USPTO.

Point of Contact:
Toby Port
Technical Info. Specialist
CM16A04
703-308-3534

*None have a
100% match.
No w-f or seq. match
except for Seq ID 5 alone.*

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:49:26 : Search time 68.4267 Seconds
(without alignments)
1483.879 Million cell updates/sec

Title: us-09-917-378-1

Perfect score: 4122

Sequence: 1 MGLVRPARAFVATAAGTAV.....GFNGSYSGTNTAPTTLCTAS 762

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	841.5	20.4	666	19 AAW34992	Thermotoga neapolitana
2	813	19.7	680	18 AAW34564	Thermotoga maritima
3	813	19.7	680	19 AAW35005	Thermotoga maritima
4	812	19.7	663	19 AAW49872	Thermotoga maritima
5	806	19.6	680	19 AAW49868	Thermotoga maritima
6	698.5	16.9	562	21 AAY69508	Acidothermus cellulosus
7	696.5	16.9	562	23 AAY79549	A. cellulolyticus
8	676.5	16.4	521	22 AAB48786	Acidothermus cellulosus
9	674.5	16.4	521	17 AAB89927	A. cellulolyticus
10	670.5	16.3	521	22 AAB48787	Acidothermus cellulosus

11	668.5	16.2	521	22 AAB48788	Acidothermus cellulosus
12	624	15.1	1751	20 AAV13493	Truncated cellulase
13	624	15.1	1751	23 AAE16324	Active cellulase p
14	589.5	14.3	1426	23 AAY13492	Truncated cellulase p
15	589.5	14.3	1426	23 AAE16323	Active cellulase p
16	552	13.4	431	21 AAG27501	Arabidopsis thaliana
17	552	13.4	431	21 AAG42172	Arabidopsis thaliana
18	552	13.4	431	23 ABB93972	Herbicidally active
19	552	13.4	442	21 AAG42171	Arabidopsis thaliana
20	552	13.4	443	21 AAG27500	Arabidopsis thaliana
21	544.5	13.2	416	22 AAG79297	Amino acid sequence
22	541	13.1	408	23 ABB92275	Herbicidally active
23	526.5	12.8	431	23 ABB90809	Herbicidally active
24	519	12.6	414	23 ABB92274	Herbicidally active
25	498	12.1	448	23 ABB93325	Herbicidally active
26	495	12.0	431	23 ABB93333	Herbicidally active
27	494	12.0	427	21 AAY93441	Amino acid sequence
28	488	11.8	551	18 AAW18790	Corrected bacillus
29	484	11.7	1853	19 AAW43108	C. thermocellum ce
30	479	11.6	531	18 AAW15238	Scalofidin protein
31	477	11.6	375	21 AAG27502	Arabidopsis thaliana
32	477	11.6	375	21 AAG42173	Arabidopsis thaliana
33	472	11.5	351	23 ABB92490	Herbicidally active
34	469.5	11.4	616	20 AAY13494	Truncated cellulase
35	469.5	11.4	616	23 AAE16325	Active cellulase h
36	461.5	11.2	415	22 AAE00417	Lycopodium escul
37	451.5	11.0	499	14 AAR42122	NR-1 cellulase. h
38	448.5	10.9	782	12 AAY15625	Cellulase AE-1. A
39	446.5	10.8	476	21 AAY54123	A mammalian linker
40	439.5	10.7	677	23 AAY98063	Bacillus subtilis
41	438	10.6	700	12 AAR13227	Novel endoglucanase
42	426.5	10.3	369	20 AAV14150	Protein encoded by
43	415.5	10.1	403	23 ABB91820	Herbicidally active
44	415.5	10.1	493	20 AAY28850	Pectate lyase-link
45	415.5	10.1	493	20 AAY43218	Pectate lyase CMO

ALIGNMENTS

RESULT 1
AAW34992
ID AAW34992 standard; Protein; 666 AA.
XX
AC AAW34992;
XX
DT 21-MAY-1998 (first entry)
XX
DE Thermotoga neapolitana endoglucanase.
XX
DE Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase.
XX
OS Thermotoga neapolitana (Clone 56Gp1).
XX
PN WO9744361-A1.
XX
PD 27-NOV-1997.
XX
PF 22-MAY-1997; 97WO-US08793.
XX
PR 22-MAY-1996; 96US-0651572.
XX
PA (RECO-) RECOMBINANT BIOTOCALYSIS INC.
XX
P1 Lam DE, Mathur EJ;
XX
DR WPI: 1998-018435/02.
DR N-PSDB; AAT94200.
XX
PT Endoglucanase(s), preferably from archaeal bacterium, AEP11 la -
useful to degrade carboxymethylcellulose and hydrolyse of

Wed Nov 13 17:38:04 2002

KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase.

OS Thermotoga maritima strain MSH8 (Clone 6GP2).

XX WO9744361-A1.

XX PD 27-NOV-1997.

XX PF 22-MAY-1997; 97WO-US08793.

XX PH 22-MAY-1996; 96US-0651572.

XX PA (KCCO-) RECOMBINANT BIOCATALYSTS INC.

XX PI Lam DE, Mathur EJ;

XX DR WPI: 1998-018435/02.

XX DR N-PSDB: AAT94213.

XX Endoglucanase(s), preferably form archaeal bacterium, AEP111a -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose

XX PS Claim 1; Fig 10; 164pp; English.

XX This protein comprises an endoglucanase of Thermotoga maritima MSH8
CC (clone 6GP2) that is capable of degrading carboxymethylcellulose
CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It
CC has homology to an endoglucanase of archaeobacterium AEP111a (see
CC AAM34985). It can be produced from native cells or from recombinant
CC host cells, especially prokaryotic host cells transformed with a
CC plasmid or virus-derived vector including the endoglucanase DNA
CC (see AAT94213). 24 endoglucanases (see AAM34986-W35008) are claimed.
CC They can be used to degrade cellulose for the conversion of plant
CC biomass into fuels and chemicals, for use in detergents, textiles,
CC animal feed, waste treatment, and in the fruit juice and brewing
CC industries for the clarification and extraction of juices.

XX SQ Sequence 680 AA;

Query Match 19.7%; Score 813; DB 19; Length 680;

Best Local Similarity 41.5%; Pred. No. 7.9e-39;

Matches 173; Conservative 65; Mismatches 135; Indels 44; Gaps 13;

OY 19 AVAAATIGSTIMPSATAAPACFTVTCGQFVLNCLPYRYGCTNNYXLSYOSHADVDDVL 78

DB 12 SVSAEFLLLIVELSVLFASDEFYKVENGRKFNALNKEFRFGSNYYMHYKSNMGIDSVL 71

OY 79 AKAAAMNLSVITWCFIDIGSLDGSVPTIDGNKNGFYQYWDPSGAPAYNDG----PTG 134

DB 72 ESARDMGKIVLRWGF-----LDGESYCRDN-----TYMHPEPGVFGVPEGISNAQSG 120

OY 135 LQGLDYATASAAHGLRVIVVLTNDWKEFGMDQYDKWYGLPYHDNFTDPTQOAYKNW 194

DB 121 PERLDYTVAKAKELGKLVIVLVNWDFFGMNQYVRMFGGTHDDFYRDEKIEKYKY 180

OY 195 VNHLNRVNSITGVYTKNDPTIFAMELANEPCVSGSLPTSGTCTOATIVNWDMSAY 254

DB 181 VSFLNVHNTVTVGVYREPTIMAMELANEPC-----ETDKSGN-----TLVWVKEMESY 232

OY 255 VKSIDPNHMSVDEGCF---YIGSTQSGC---WPYNDPSGDVNDNALLRVKNIDFCTYHL 308

DB 233 IKSLDPNHLVAVGDEGCFYSNVEGFPYCGEAWYNGWS-GVDWKKLLSITVDFCTEHL 291

OY 309 YPNYWGQN-----ADNGTOWIKDHIANAAIGKPTILEEFG--WQTP-DRDSVYQTWTQV 361

DB 292 YPSHWGVSPENYAQWAKWIEDHIIKIAEIGKPVVLEEYIPKSAFVNRTALIRLWNLV 351

OY 362 RTNGEAGNFWNLACNVACP-----YPNYDGFNVYHPSSTATVLAASE-ALAIPTG 411

DB 352 YDLGGDGAMFWNLGIGESDRDERGYYPDYDGFRIVNDSDPEAEILREYAKLFNTG 408

RESULT 4

AAW49872

ID AAW49872 standard; Protein: 663 AA.

XX AC AAW49872;

XX DT 21-DEC-1998 (first entry)

XX Thermotoga maritima MSH8-6GP2 glycosidase.

XX Glycosidase; MSH8-6GP2; thermostable enzyme; oligosaccharide;
KW glucose; sugar; baking; textile; detergent; beta-galactosidase.

XX OS Thermotoga maritima strain MSH8-6GP2.

XX PN WO9824799-A1.

XX PD 11-JUN-1998.

XX PF 08-DEC-1997; 97WO-US22623.

XX PR 10-OCT-1997; 97US-0949026.

XX PR 06-DEC-1996; 96US-0056916.

XX PA (DIVE-) DIVERSA CORP.

XX PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;

XX DR WPI: 1998-362407/31.

XX DR N-PSDB: AAV36921.

XX Glycosidase enzymes from organisms of the genera Staphylothermus,
PT Pyrococcus and Thermococcus - for deriving sugar from
PT oligosaccharides, useful in the e.g. food processing, textile or
PT baking industries

XX PS Claim 1; Fig 15a-d; 92pp; English.

XX This is the amino acid sequence of glycosidase MSH8-6GP2, deduced
CC from a polynucleotide (see AAV36921) of clone 6GP2 of Thermotoga
CC maritima MSH8 that grows optimally at 85 degC in high salt medium.
CC The invention provides 18 polynucleotides (see AAV36907-24) coding
CC for thermostable glycosidases (see AAW49858-75) having glucosidase,
CC alpha-galactosidase, beta-galactosidase, beta-mannosidase,
CC beta-mannanase, endoglucanase or pullulanase activity. Vectors and
CC host cells are also claimed. A method is provided for producing
CC the enzymes by recombinant techniques. A claimed method for
CC generating glucose from soluble cell oligosaccharides comprises
CC contacting a sample (selected from dairy products, fruit juices,
CC detergent, textile, guar gum, animal feed, plant biomass or waste
CC product) containing oligosaccharides (selected from maltose,
CC cellobiose, lactose, sucrose, raffinose, stachyose, verbascose,
CC cellulose, starch, amylose, glycogen, disaccharides, polysaccharides
CC and pullulan) with one of the claimed glycosidases such that glucose
CC is produced.

XX SQ Sequence 663 AA;

Query Match 19.7%; Score 812; DB 19; Length 663;

Best Local Similarity 42.8%; Pred. No. 8.8e-39;

Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

OY 41 FVTASGQFVLNGLPYRYGCTNNYXLSYOSHADVDDVLAKAAMNLSVITWGFIDIGSL 100

DB 17 FVKVENGKFNALNKEFRFGSNYYMHYKSNMGIDSVDLESARDMGKIVLRWGF-----L 71

OY 101 DGSVPTIDGNKNGFYQYWDPSGAPAYNDG-----PTGLOGLDYATASAAHGLRVIVVL 156

DB 72 DGESYCRDN-----TYMHPEPGVFGVPEGISNAQSGFERLDYTVAKAKELGKLVIVL 125

OY 157 TNDWKEFGMDQYDKWYGLPYHDNFTDPTQOAYKNMNVNHLNRVNSITGVYTKNDPTI 216

[illegible]

RESULT	9
AAK89927	
ID	AAK89927 standard; Protein; 521 AA.
XX	
AC	AAK89927;
XX	
DT	08-OCT-1996 (first entry)
XX	
DE	A. cellulolyticus E1 endoglucanase.
XX	
KW	El endoglucanase; cellulase; cellulose; saccharification; ethanol.
XX	
OS	Acidothermus cellulolyticus.
XX	
PH	Key Location/Qualifiers
FT	Active-site
Domain	1..363
FT	/label= Catalytic-domain
FT	364..417
FT	/label= Linker
FT	/note= "proline/serine/threonine-rich linker domain common to multi-domain microbial cellulases"
FT	

FT	Domain	418..521	
FT	/label=	Cellulose-binding_domain	
XX			
PN	WO9602551-A1.		
XX			
PD	01-FEB-1996.		
XX			
PF	14-JUL-1995;	95WO-US08868.	
XX			
PR	15-JUL-1994;	94US-0276213.	
PR	26-SEP-1989;	89US-0412434.	
PR	27-JAN-1992;	92US-0826089.	
PR	21-SEP-1993;	93US-0125115.	
XX			
PA	(WIDE) MIDWEST RES INST.		
XX			
PI	Himmel ME, Laymon RA, Thomas SR;		
XX			
DR	WPI: 1996-105843/11.		
DR	N-PSDB; AAT12337.		
XX			
PT	New Isolated DNA encoding endoglucanase - obtld from Acidothermus		
PT	cellulolyticus, used for prodn of the enzyme for use in cellulose		
PT	hydrolysis.		
XX			
PS	Claim 1; Page 22; 34pp; English.		
XX			
CC	Acidothermus cellulolyticus El endoglucanase (AAR89927) is useful		
CC	for hydrolysing cellulosic biomass to sugars for simultaneous or		
CC	subsequent fermentation to ethanol. It shows optimal activity at		
CC	83 deg C. The amino acid sequence of the mature enzyme was deduced		
CC	from an El endoglucanase gene (AAT12337) cloned from A. cellulolyticus		
CC	Putative signal peptides were also identified (see also AAR89928-29).		
CC	Cloning of this gene allows large-scale, low-cost prodn. of		
CC	recombinant El endoglucanase, using pref. Saccharomyces, Zygomonas		
CC	or E.coli hosts.		
XX			
SO	Sequence	521 AA:	

```

Query Match      16.4%:  Score 674.5;  DB 17;  Length 521;
Best Local Similarity 27.2%:  Pred. No. 5.5e-31;
Matches 210;  Conservative 88;  Mismatches 177;  Indels 297;  Gaps 32;

Qy      37  APAGFVTASGGQFV--LNGLPYRYGCTN-----NYILSYOSHADVDDVLAKAQAAMNLSV 88
Db      1  AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYSKMLDQIISLGYNV 60

Qy      89  IRTWGFIDIGSLDGSVPTIDGNKNGFYFYQWDPSTGAPAYNDGPTGLGGLDYATASAAAH 148
Db      61  IRL-PYSDDILKPGTMP---NSINFYQMNQD-----LQGLTSLQWMDKIVAYAGQI 107

Qy     149  GLRVIVLNTDWEKFGGMDQYDKWGLPYHDNFYDPTQOAYKNWVHLLNRVNSITCV 208
Db     108  GLRII--LDRHRPCDSC--QSALWY-----TSSVSEATWISDL-----QALAQ 146

Qy     209  TYKNDPTPTIFAWELANEPR---CVSGTLPTSGCTQATIVNW---VDQMSAYVKSIDPNH 262
Db     147  RYKGNPTVVGDLHNEPHDPACWCCGD-PS-----IDHRLAAERAGNAVLSVNPNL 196

Qy     263  MYSV-----GDEGFYTGSTQSG--WPYNPDPSCVDNALLRVKN-----IDRGTVHL 308
Db     197  LIFVEGVQSYNGDSYWMGGLQAGQYF-----VVLNVPNRLVYSAHDYAT-SV 244

Qy     309  YPNYQGNAD-----WCTQW---IKDHIANAAAIKPTILEFFGHQTPDRDSVYQTW 357
Db     245  YPQTFSDPTTPNNMPGLWNKMGVLFNONTA-----PWLGEFG---TTLQSTQDTW 295

Qy     358  TOTVRTNCEAGNFWMLAGNVNGQYPYNYDGFNVYPSSTATVLAASEALISTGTSPPPS 417
Db     296  LKTL-----L----- 299

Qy     418  PSSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 477
Db     418  PSSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 477

```


XX DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y245G.
XX KW E1 endoglucanase; glycosyl hydrolase; insoluble substrate;
XX KW cellulose hydrolysis; ethanol production; fermentation;
XX KW mutant; mutein.
XX KW Acidothermus cellulolyticus.
OS Synthetic.
XX PN WO200070031-A1.
XX PD 23-NOV-2000.
XX 19-MAY-2000; 2000WO-US13971.
XX 19-MAY-1999; 99US-0134925.
XX (MIDE) MIDWEST RES INST.
XX Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
PI Decker SR;
PI WPI; 2001-061226/07.
DR Preparation of glycosyl hydrolase with an increased catalytic activity
on insoluble substrate
Claim 5; Page 27-29; 30pp; English.
XX The invention relates to a method for making glycosyl hydrolase mutants
with increased catalytic activity with either insoluble or soluble
cellulose substrates relative to the wild-type enzyme. The method for
making a glycosyl hydrolase with increased insoluble substrate catalytic
activity comprises replacing an active site-associated amino glycosyl-
stabilising amino acid with an amino acid that does not strongly bind a
disaccharide product in the active site. Conversely, the method for
making a glycosyl hydrolase with increased soluble substrate catalytic
activity comprises replacing a hydrophobic substrate-binding amino acid
with a positively charged residue. The invention also discloses mutants
of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)
produced according to the method of the invention. The Y245G mutant
(AAB48788) has improved activity with insoluble substrates, and the W42R
(AAB48786) and Y82R (AAB48787) mutants have improved activity with
soluble substrates. The invention also encompasses DNA encoding these
mutants. The glycosyl hydrolases of the invention are used as catalysts
for cellulose hydrolysis to produce sugars that can be fermented to
produce fuels such as ethanol. The present sequence represents the
Acidothermus cellulolyticus E1 endoglucanase Y245G mutant.
XX
SQ Sequence 521 AA;
Query Match 16.2%; Score 668.5; DB 22; Length 521;
Best Local Similarity 27.1%; Pred. No. 1.2e-30;
Matches 210; Conservative 87; Mismatches 176; Indels 301; Gaps 31;
QY 37 APAGFYTAGGQGVV-LNGLPYRYCGTN-----NYLSYQSHADVDVLAQAAMNLVS 88
Db 1 AGCGYHRTSGREILDANNVPVRIAGINWFGFETCNVWHGLWSRDYRSLMDQIKSLGYNT 60
QY 89 IRTWGFIDIGSLDGSVPTIDGNKNGFYFQYWDSTGAPAYNDGPTGLQGLDYATASAAH 148
Db 61 IRL-PYSDDILKPTMP-----NSINFQMNQD-----LOGLTSLQVMDKIVAVAGQI 107
QY 149 GLKRVIVLTMNDKFEFGMDQYDKWGLPYHDNFYDPTQQAYKNVHLLNRVNSITGV 208
Db 108 GLRII--LDRHRPDCSG--QSALWY-----TSSVSEATWISDL-----QALAQ 146
QY 209 TYKNDPTIFAWELANPR---CVSGTLPTSGTCTQATVNW---VDQMSAYVKSIDPNH 262
Db 147 RYKGNPTVVGFDLRNEPHDPACGCCD-PS-----IDWRLAAERAGNAGNVLSPNPL 196
QY 263 MVSV-----GDEGYIGSTQSG-WPYNDPSGDGVNALLRVKNIDFGTYHLY----- 309

Db 197 LIFVEGVQSYNGDSYWMGNGLOGAGQYP-----VVLNVPNRLVYSAHDYATSVG 245
QY 310 -----PNYWGQNADMGCTOWIKDHIANAAAGKPTLLEFGMGWTPDRDSYVQ 355
Db 246 PQTWFSDPPTFPNNMPGIWNKN--WG-YLFNQNJIA-----PVMLGEGF--TTLQSTTDQ 293
QY 356 TWTOTVRTNCEAGNFWMLAGNVNQCYPNYDGFNVYPPSSSTATVLAASEALAIATGTSHP 415
Db 294 TWLKTLL-----VQY----- 299
QY 416 PPSSSSPSSP 475
Db 300 -----VQY-----L 304
QY 476 KPGQLVNTGSSVDLSTVTVRYVWFTTRDGGSSTLVYNCDMAAMCGNIRASFGSVNPAIP 535
Db 304 RP-----TAQY-----CADSFQWTFWSNPD- 325
QY 536 TADTYLQLSFTGGTLAAGSGTGEIONRVNKSMDSNFDETDNDYSYCTNTAFQDMTKVTYV 595
Db 326 -----GDTGGIL-----KDDMTQVDTVKD-----GYLAPIKSSIP- 355
QY 596 NGRLVMGTEPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTP 649
Db 356 -----DPVGASASPSQSP 407
QY 650 PPSPPSPS-VSSSGVCGCRATYVNVNSDMGSGTATVTVTNTGSRATSGMTVAMSGCQNV 70H
Db 408 PTASPTSPSTAASCARCTASYQVNSDMGNGFTVAVAVTNSGVSATKTWTVSWTEGNOVI 467
QY 709 TNYWNTALTQSGASVATNLNNVNIQPCQSTTFGFCNGSYSGTNTAPTTLCTAS 762
Db 468 TNSWNAAVTONGOSVTAARNNSYNIQPCQNTTFGFOASYTGSNAAPTVAACAAS 521
RESULT 12
AAY13493
ID AAY13493 standard; Protein; 1751 AA.
XX AC AAY13493;
XX 30-JUL-1999 (first entry)
XX Truncated cellulases comprising amino acid sequence.
XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
XX Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
XX cotton-containing fabric; stonewashing.
XX Unidentified.
XX EP921188-A2.
XX 09-JUN-1999.
XX 15-SEP-1998; 98EP-0810919.
XX 19-SEP-1997; 97US-0932571.
XX (CLRN) CLARIANT FINANCE BV1 LTD.
XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX N-PSDB; AAX55662.
DR WPI; 1999-315403/27.
XX New truncated cellulase proteins, useful in detergents and for
producing stonewashed denim
PS Claim 7; Page 37-41; 65pp; English.
XX

The invention relates to a recombinant cellulase active protein free of proteinases of native thermophilic and alkaliphilic origin, comprising the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3, Cel 6 or Cel E3/B5, or a stability region from one of the defined full-length sequences, or functional equivalents. Cel B5 extends from amino acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino acid K635 to N1426 in the sequence shown in AA131492; Cel E1 extends from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V123 to K1751 and the stability region extends from amino acid E482 to G635 in the sequence shown in AA131493; Cel E3/B5 is shown in AA131494. The new enzymes are useful in laundry detergent compositions to prevent or remove staining, backstaining or graying, for use on cellulosic materials including cotton-containing fabrics. They are especially useful for preventing redeposition of colorant during stonewashing, and for processing of textiles where cellulose breakdown is required. The new truncated enzymes show reduced redeposition of dye compared to using non-truncated cellulase compositions.

Sequence	1751	AA:
5Q		

Query Match 15.1%; Score 624; DA 20; Length 1751;
Best Local Similarity 25.6%; Pred. No. 1.8e-27;
Matches 234; Conservative 140; Mismatches 278; Indels 262

QY	7	PARAFVA--TAAGTAVANAAMTLGSIWPSA-----TAAPAGFVTASG-----46
Db	183	PASSVAAETAAASLAVASVVIKERNSSOKAASVQLQHAQDLFEFADTTTSDGADYTAATGFYTS 242
QY	47	GQFVLN---GLPYRYGCTNNYILSYSGHADVDVLAKAAMNLSVTRTNGCFIDIGSL---100
Db	243	GGFIDDLGWAAVWLXIATND--SSYLTKA--EELMSEYANGTWITQCHDDVRYGTLIML 298
QY	101	-----DCSVPTIDCNKNGHYFQW--DPSTCAPAYNDGPTGIGLDYAIASAAAHGL 150
Db	299	AKITGKELYKGAV-----ERN-----LDHWTDRITPTPKGMAYLTGWSLRYATTAA-----345
QY	151	KVIVLVLNDWKEFGGMDYDKW-----YGLPYHDNFYTDPRTOQA 190
Db	346	-FIACVYADN---SGCDSNKTKYLNFAKSOIDYALGSGTGRSPVVGFGNYPOHPIHINA 401
QY	191	YKNWVN-----HILLNRVNSITG-----VTYKNDPTIFAWELANEPRC-VGSCLTPT 235
Db	402	ISSWANSMKIPEYIIRHIL--YGALVGGPGSDSYNDIDTY-----VQNEVACDYNAGIVGA 456
QY	236	SGTCTQATIVNWWDMSAVKSIUPNHVSVGDEGF-----IGSTOG-----S 279
Db	457	LAKWQLYGCEPIDDF-----KALE-----TPTNDEIFVESKFGNSQGPNTVEVLSYLYNRT 508
QY	280	GWP-----YNDPSPGVDDNALLRVKNTIDFGTYIILY-----PNYWGQNAD-----318
Db	509	GWPVRVTDKLSFKFYFIDLTETJQAGYSPDVVKVD--TYVIEGCKISGPYWDKNRNIYVY 566
QY	319	---WGTOWIK---DIIHANAAGIPTILLEEFCHWOTPDSDSYQWTQTIVRTNGEAGWN 370
Db	567	LVDMSGTKIYPGGEVEIHKQAF--KISVPGQYPMW-PTNDPSYKGLTSOLEKN-----617
QY	371	FWMLAGNVNQOPYNYDGFNVVYPSSTATVLAASEALAISTGTSPPPSPSSSPSSSPSPSP 430
Db	618	-----KYIAAYDNNNLVW-----GLEPGAATSTP-----APTSTPTTP 651
QY	431	SPSPASPSASPSASSPSPPSSSPVSG-GVKVOYKKNDSAPGDNQIKPGLQLVNTGSSSV 489
Db	652	TPITPTVATP--TPITPTPTPGSGVGVKLYKNNETSASTGSLRPWFKIVNGCGSSV 709
QY	490	DLSTVTYVWFTRDGGSSTLVYKCDWAAMCGGNIRASFGSVNPAITPATDYTLQLSFTGCT 549
Db	710	DLSRVKIRYWTVDGDKPOSAB--CDWAIGASNVTFENFKLSGCVSGADYILEVGFSSGA 768
QY	550	--LAACGSSTGEIQNRNVKNSDWSNFDETNDS-----YGTNTAFODWTKVTYVNGRLV 600
Db	769	GQLQPKDQTDIGQVRNKNDNSYNGQADMWLSQNTNGEN-----AKVTLYVDGVVLV 822

CC The present sequence is Cele protein.

```
XX SQ Sequence 1751 AA;
Query Match 15.1%; Score 624; DB 23; Length 1751;
Best Local Similarity 25.6%; Pred. No. 1.4e-27;
Matches 234; Conservative 140; Mismatches 278; Indels 262; Gaps 46;

QY 7 PARAFVA-TAAGTAVAAATLGSITMPSA-----TAAPAGFTVTSAG--- 46
DB 183 PASSVVAETAASLAVASVVLKERNQKAAASVQLQAKDLFEADTTRSDAGTAAATGYTS 242
QY 47 GQFVLN---GLPVRGNTNNYLYSYOSHADVDDVLAKAAMNLSVIRTWGFDIGSL--- 100
DB 243 GGFIDDLGMAAVWLYIATND--SSYLTKA--EELMSEYANGTNTWTQCDVDRYGTLM 298
QY 101 -----DGSVPTIDGKNGFYFYQW-DPSTCAPAYNDGPTCLQGLDYAIAASAAHGL 150
DB 299 AKITGELYKGAV-----ERN---LDHMTORITTPKGMAYLTGWSGLRATTAA----- 345
QY 151 RVIVLTNDMKFCGMDQYDKW-----YGLPYHDNFYTDPRTOQA 190
DB 346 -FLACVYADW---SGCDSSNKKTKYLNAFAKSIDYALGCTCRSFVVGCTNYPOHPHHRNA 401
QY 191 YKMWVN-----HLLNRVNSITG-----VTYKNDPTIFAWELANEPRC-VGSGTLPT 235
DB 402 HSSWANSMKIPEYHRHIL--YCALVGCGPGSDSDSYNDIDTDY---VQNEVACDYNAGIVGA 456
QY 236 SGCTCTQATIVNWDQMSAYVKSIDPNHMSVGDGEFY---IGSTQG-----S 279
DB 457 LAKMYOLYGEPIDDF-----KALE-----TPTNDEIFVESKFGNSQSGFTEVYSIYNRT 508
QY 280 GWP-----YNDPSGDVNNALLRVKNIIDFGTYHLY-----PNYWGQAD----- 318
DB 509 GWPFRVTDKLSFYKFDLTLELIQAGYSPDVVKVD--TYTIEGCKLSCPYVWDRNRNIYV 566
QY 319 ----WGTOWIK-----DHIAANAAIGPRTILEEFCWQTPDRDSVYQVWTQVTRTNGEAGWN 370
DB 567 LVDFSTCKIYPGGEVEHKKQAOF--KISVPQGYPWD--PTNDPSYKGLTSOLEKN----- 617
QY 371 FWMLAGNVNGQPNYDGNFVYVPSSTATVLAASEALAIAGTSPPPSPSSPSSSPSSPSP 430
DB 618 -----KYTAAADNNNLVW-----GLEPGNATSTP-----APTSTPTPTP 651
QY 431 SPASPSASPSASSPSPSPSPSPVSG-GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSSV 489
DB 652 TPTPTVTATP--TPTPTPTPTGPGTSGVKVLYKNNETSASTGSIKPMFKIYVNGSSSV 709
QY 490 DLSTVTYRWFTRDGGSTLVYNCDAAMCCGNIRASFGSVNPTATPTADTYLQSLFTGGT 549
DB 710 DLSRVKIRIYWTVDGDKPOSAY-CDWAQICASNVTNFVKLSGVSAGADYLYEVGFSSGA 768
QY 550 --LAAGGSTGEIQNRVYNKSDWSNFDETNDYS-----YCTNTAFQDWTKVTVYVNGRLV 600
DB 769 GQLOPKGDKTGDIOVRFNKKNDSYNNQADMSWLMQNTNYGEN-----AKVLYVDGLV 822
QY 601 WGTERTSPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSP 654
DB 823 WQEPFGGATPAPTSTATPTPTATPTPTP--TPTPTPTVSATPTPTAPTASPVGGSYWT 880
QY 655 SPVSSSSGVGRATVYVNSDWGSG---FTATVTVTNGS-----RAT 693
DB 881 SES-----YGALKVYVANGLSPTNLPKIKIENVGTAVDLSRVKRYWYITIDGEAT 935
QY 694 SGWTVAMS-----FCGNTQTVTNYMNTALTOSCATVATNLSNNV---LQPGQ 738
DB 936 QSVSVASSINPAYIDVKFVKLCANAGGADYVEIGFKSGAGVLAAGQSTKEIRLSIQK- 994
QY 739 STTFCFNGSYSGTN 752
DB 995 -----SGSYNQSN 1002
```

RESULT 14

```
AAV13492
ID AAY13492 standard; Protein; 1426 AA.
XX
AC AAY13492;
DT 30-JUL-1999 (first entry)
XX
DE Truncated cellulases Cel B4/5 and Cel B5.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
OS Unidentified.
XX
PN EP921188-A2.
XX
PD 09-JUN-1999.
XX
PF 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
PA (CLRN ) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI; 1999-315403/27.
DR N-PSDB; AAX55661.
XX
PT New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
XX
PS Claim 7; Page 34-37; 65pp; English.
XX
CC The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
CC from amino acid Y39 to D481; Cel E1/2 extends from Y39 to G635; Cel
CC E1/2/3 extends from Y39 to G812; Cel E6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or
CC remove staining, backstaining or graying, for use on cellulosic
CC materials including cotton-containing fabrics. They are especially useful
CC for preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using
CC non-truncated cellulase compositions.
XX
SQ Sequence 1426 AA;
Query Match 14.3%; Score 589.5; DB 20; Length 1426;
Best Local Similarity 24.6%; Pred. No. 1.4e-25;
Matches 199; Conservative 138; Mismatches 251; Indels 221; Gaps 33;

QY 84 MNLVSIVRTWCFDI---GSLDGSVPIT-DGNKNGFYFOYWDPS-----" 123
DB 19 MSLISLGSMSYFPVKTEAAPDWSIPSLWESYKNDFKIGVAIPARCLNSDNDKQMLKHN 78
QY 124 GAPAYND-----GPTGLQGLDYAIAASAAH-----GLRVIVL-----TNDWKE 162
DB 79 SITAENKMPESLLAGOTS-TGLSYRFSTADTTFVNTANTNNIGRHTLWHHNOTPDW-- 135
QY 163 FCGMDQYDKWGLPYHDNFYTDPRTOQAYKNMNVNHLNRYNS-ITGVYTKNDPTIFAWEL 221
DB 136 -----FFRDSGQMLSK---DALLARLKQVIYDVGVGRYKCKVYADV 174
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:50:41 : Search time 35.5819 Seconds
(without alignments)
2058.756 Million cell updates/sec

Title: US-09-917-378-1
Perfect score: 4122
Sequence: 1 MCLVRRPARAFVATAAGTAV.....GFNGSYSGTNTAPTLCTAS 762

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues 283224
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	811	19.7	669	2 D72278	endo-1,4-beta-mann
2	676.5	16.4	1331	2 A48954	mannan endo-1,4-be
3	661.5	16.0	1711	2 T31337	1,4-beta-glucanase
4	632.5	15.3	1742	2 T17120	cellulase (EC 3.2.
5	609	14.8	1779	2 T31085	xylanase - Caldice
6	603.5	14.6	1039	2 S02711	cellulase (EC 3.2.
7	526.5	12.8	411	2 D86153	hypothetical prote
8	498	12.1	448	2 T48214	endo-1,4-beta-mann
9	495	12.0	431	2 T09048	probable mannan en
10	484	11.7	1854	2 S36859	c1pA protein - Clo
11	460	11.2	508	2 G69593	cellulase (EC 3.2.
12	454	11.0	508	2 A26874	cellulase (EC 3.2.
13	451.5	11.0	499	2 A27198	cellulase (EC 3.2.
14	451	10.9	504	2 S54744	cellulase (EC 3.2.
15	451	10.9	747	2 B47093	cellulase (EC 3.2.
16	447	10.8	499	2 JN0111	cellulase (EC 3.2.
17	445	10.8	700	2 B41897	cellulase (EC 3.2.
18	426.5	10.3	369	2 T04323	mannan endo-1,4-be
19	426.5	10.3	505	2 S39962	endoglucanase - Er
20	423.5	10.3	442	2 F87348	mannanase, probabl
21	415.5	10.1	403	2 A84592	(1-4)-beta-mannan
22	412.5	10.0	915	2 A43802	cellulase (EC 3.2.
23	402	9.8	486	2 I40548	bifunctional cellu
24	394	9.6	145	2 A41897	cellulase homolog
25	380	9.2	441	2 T12011	cellulase (EC 3.2.
26	365.5	8.9	890	2 T35237	probable secreted
27	362	8.8	962	2 S03818	carboxymethylcellu
28	361.5	8.8	473	2 S50755	hypothetical prote
29	353.5	8.6	570	2 S56132	cellulase (EC 3.2.

30	335	8.1	1045	2 A39199	endoglucanase (E
31	330	8.0	872	2 S49541	cellulase - Cellul
32	313.5	7.6	1090	2 S59077	cellulose 1,4-beta
33	312.5	7.6	586	2 PC6006	scaffolding protei
34	311.5	7.6	484	1 A24994	cellulose 1,4-beta
35	311.5	7.6	1162	2 T30433	scaffolding protei
36	310	7.5	611	2 S76211	hypothetical prote
37	308.5	7.5	986	2 S12021	thermoactive cellu
38	308	7.5	1483	2 C97012	probably celluloso
39	305.5	7.4	1230	2 S47466	cellulose 1,4-beta
40	301.5	7.3	579	2 T35240	probable secreted
41	300	7.3	279	2 T10361	hypothetical prote
42	298	7.2	879	2 A47704	endoglucanase 1 (E
43	294.5	7.1	2232	2 T34434	hypothetical prote
44	285.5	6.9	474	2 S15921	protein TPX-VT3 -
45	285.5	6.9	1032	2 T34433	hypothetical prote

ALIGNMENTS

RESULT 1

D72278

endo-1,4-beta-mannosidase - Thermomoga maritima (strain MSB8)

C:Species: Thermomoga maritima

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: D72278

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Hall, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Colton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: D72278

A:Molecule type: DNA

A:Residues: 1-669 <ARN>

A:Cross-references: GB:AE001779; GB:AE000512; NID:q4981777; PIDN:AAD36302.1; PID:q498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1227

Query Match 19.7% Score 811; DB 2; Length 669;
Best Local Similarity 42.8%; Pred. No. 2e-33;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

Qy 41 FVTASGGQFVLNGLPYRYGGTNNYVLSYQSHADVDVLLAKAQAAMNLSVIRTWGFDIGSL 100

Db 23 FVKVENGKFKALNGKEFRFIGNSNYYMHYKSNRMIDSVLESARDMGKIKVLRWGF-----L 77

Qy 101 DGSVPTIDGNKNGPYFOYMDPSTCAPAYNDG-----PTGLQGLDYAIAASAAHGLRVIVVL 156

Db 78 DGESYCRDKN-----TYMHPEPGVFGVPEGISNAQSGFERLDYTVAKAKELGIKLVIVL 131

Qy 157 TNDKKEGGMDQYDKWYGLPYHDNFYDTPRTOQAYKKNVHLLNRVNSITGVTKNDPTI 216

Db 132 VNNWDDFGGMNQYVRWFGGTHHDDFYRDEKIKEEYKYVSFLVNHVNTYTGVPYREEPTI 191

Qy 217 FAWELANEPKRCVSGSGLTPTSGCTCTOATIVNWDQMSAYVKSIDPNHMYSVGDEGF---Y 273

Db 192 MAWELANEPKRC---ETDKSGN-----TLVENVKEMSSYIKSLDPNHLVAVGDEGFFSWE 243

Qy 274 GSTGSG---WPYNDPDSGVNDNALLRVKNTDFCTYHLYPNYWGQN-----ADWGQTQIKD 326

Db 244 GFKPYGGEAWAYNGWS-GVDWKLLSTETVDFCTFHLYPHSHWGVSPENYAQWAKWIED 302

Qy 327 HIAAANAIGKPTILEEFG--WQTP-DRDSVYQTVTQTVRTNGEAGHWNFMLAGNVNCP- 382

Db 303 HFKTAKKIGKPVLYLEEYGIKPSAPVNRATYRLWNLNDLYDLGGDGMFWMLAGICEGSDR 362

Qy 383 -----YPNYGDNVYYPSSSTATVLASE-ALAIISG 411

Db 363 DERGYYFDYDGFRIVNDSDPEALIREYAKLFNTG 397

Db	577	SNDSYFDIKG--VSSGSVVKTKYI-----PLYGDGVKVGCDGPCTSGAT-----	619
Qy	376	GNVNGQYPNVDGNVYVPSTATAVLASEALAIStCTSPPPSPSSPSSPSsp--SPSPS	433
Db	620	-----PPTTATATPTPTPTPTPTPTPTPTSTATPTPTPTPTPTPTPTPTPTPTPT	661
Qy	434	ASPSASPASSSPSSPSPPVSCG-VKVOYKNDSAPGDNOKPCQLQLVNTGSSSDLS	492
Db	662	PTTATPT	721
Qy	493	TVTVRYFTRDGGSSLTVLYNCDMAAGCGNIRASFOSVANPATPTADTYLQLSFTTCGT--L	550
Db	722	RVTIRYWTVDGDKAQAIS-DNAOICASNVTFKFVKLSSVSAGDYYLEIGFKSGAQGL	780
Qy	551	AAGSGTGEIONRVKNSDMSNFDETNDYS-----YGTNTAFQDTKTVTIVYNGRLVNGT	603
Db	781	OAGKDTEIQIRFNKSDSWINOGNDWSMQSWNTGEN-----VKYTAYIDGLVLMVGQ	834
Qy	604	EPSCTSPTSPTSP	652
Db	835	EPSGATPTPTAPT--VTPT	892
Qy	653	-----SPSPSVSSSGVGCKRTVVVN-----SDWSGGTFATVVT-684	
Db	893	NKETNSTNTTIRPLWKVYVNTGSSSIDLSRVRTIRYWYTVDCDKAQAISDWAOIGASNVTF	952
Qy	685	--VTMTGRATSGMTVAWSFGNGTFV-----TNV-----WNTAL716	
Db	953	KFVKLSSVSACADYYLEIGFKSCAGLOLQACKOTGEIQIRFNKSDSWNTNOGNDWSMQSM	1012
Qy	717	TQSGASTVATNLSY-NNVIOPGOSTTTFGNGSYSGTNTAPTLL758	
Db	1013	TNYGENVKVT--AYIDGLVLMVGQEP-----SGATPTPTPTAT1045	
RESULT 4			
Tl7120			
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum			
C:Species: Caldocellum saccharolyticum			
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000			
C:Accession: Tl7120; A43745			
R:Te' O, V.S.; Saul, D.J.; Bergquist, P.L.			
Appl. Microbiol. Biotechnol. 43, 291-296, 1995			
A>Title: CelA, another gene coding for a multidomain cellulase from the exte			
A:Reference number: Z18698; MUID:95336703; PMID:7612247			
A:Accession: Tl7120			
A>Status: preliminary; translated from GB/EMBL/DDBJ			
A:Molecule type: DNA			
A:Residues: 1-1742 <TEO>			
A:Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PTDN:AAA91086.1			
R:Luehl, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.			
Appl. Environ. Microbiol. 57, 694-700, 1991			
A>Title: Cloning, sequence analysis, and expression in Escherichia coli of			
A:Reference number: A43745; MUID:91247819; PMID:2039230			
A:Accession: A43745			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1516-1544,'A',1546-1742 <LUES>			
A:Cross-references: EMBL:M36063; NID:g144292; PID:AAA72860.1; PID:g144293			
C:Genetics:			
A:Gene: celA			
A:Keywords: glycosidase; hydrolase			
Query Multih			
Statistical Similarity 15.3% Score 632.5; DB 2; Length 1742;			
Matches 227; Conservative 143; Mismatches 290; Indels 335; Gaps			
Qy	7	PARAFAVTAAGTAVAAATLGSIITMPSA-----TAAPAGFTVASGQGFV50	
Db	170	PGSAAVTAETAASLAASIVLKDRNPKTATYLQHAKLDFEAEVTKSDSGGYTAANG----	225
Qy	51	LNGLPRYGCTNNYILSYOSHADV-----DDVLAKAQAMNLSVITWGFD96	

Db	226	-----YNSWSCFYDELSWAAVWLYLATNDSTYLTAKB-----SYVQNN-----	264
Qy	97	IGSLDGSVPTIDGNK-----NG-----FYFQWDPST	123
Db	265	-----PKISGSNIIDYKWAHCWDDVHNGCAALLLAKITDKIYTKOIIIESHLIYWTTCY	316
Qy	124	CAPAYNDGPTCLOGLD-----YALASAAAHGLRVIIVLITNDMKFPGMIQYDKWYGLIY	177
Db	317	NGERIKYTPKGLAWLDQWGLSRLATTTAF-----LAFV-----YUSWSGZP-	357
Qy	178	HDNFYDPRTOOAYKKNVNHLLNRVNSITGVTVYKNDPTIPANELANFPCVCGSTLPIYSG	237
Db	358	-----TGKKETRYKFGESOIDYALCSTGRSF-----VWGFNTPPKR	394
Qy	238	TCTOATIVNMWDOMSAVYKSIDPNHIWYSGDEGFYIGSTQGSWPYNDPSDGVGNNAIILK	297
Db	395	PHHRTAHSSWADS-----QSTPSYRHRTL--YCALVGG-PCGSDSYTDIDISNYVNVNEAC	446
Qy	298	VKNIDF-----GTHYLY-----PNYWCNADWCTQWIKDHIANAAIG-----KPTJLIEEP	343
Db	447	DYNAGEFVICALAKMYLLYGGNPIIDFKAIETPTNDEFFVEAGINASCOTNIEIKAIIVNQOS	506
Qy	344	CW-----OTPDSDSYVQYQTTQTVKTRNCBAGWN-----FWMLA	375
Db	507	GWPARATNKLKRYFYVDLSELIKAGYSPNOLITLSTNYNOGAKYSGPYWDDSSRNIYITLY	566
Qy	376	GNVNGQYP-----NYDGFNY-----YPSSTATYLASEALIT--	408
Db	567	DFTGTLIYPGODKYKKEVOFRIAAQPONVQWDMNSDYSODIKGVSSGSVVKTKYIPLYD	626
Qy	409	-----STGCTSP-----PPSPSSPSSPSPSPSPSPSPSPSPSPSPSPSPS	451
Db	627	EDIKWGEETGCVSTPTASVTPTPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPT	686
Qy	452	SSP-----VSGVKVQYKKNDSAPGDNOIKPGLQIIVNTGSSVDLSVTVTYK	498
Db	687	STPTVTPTPTPVSTPATSGQIKVLYANKETNSTTNTIRPWLKVVNSGSSSIDLSKVTIRY	746
Qy	499	WFTDRGSSSTLVYNCDMAAGCCNIRASFGSVNPATPTADTYLQIISFTCGT--LAAGGST	556
Db	747	WYTVDGERAQAIS-DWAQICASNVTFKVKLSSSVSGADYYLEICFKSGAGQLPGKDT	805
Qy	557	GEIQNRVKNKSDWSNFDENDYSY-NTNTAFODMTKVTYVNGRLVWGTFPSCZSPSPSPS	615
Db	806	GEIQIRNKDDWSNYNGNDWSHIQSTSYGENEKTATIDGVLVWGOBPGSTTTPAPTST	865
Qy	616	P-----SPTPSP	652
Db	866	PTVTVTPT	925
Qy	653	-----SPSPSVSSSGVCCRATVYVN-----SDWGSFPTATVT-----VNTGCS	690
Db	926	IRPWLKVVNSGSSSIDLSRVTIRYTVYDGERAQAISDWAQICASNVTFFKVLSSSVS	985
Qy	691	RATSGMTVAHSFGGQNTV-----TNY-----WNTALTOSGASVTAT	726
Db	986	GADYYLEIGFKSGAGLOPKDGTGEIQIRFNKDDWSNYNGNDWSHIQSTSYGENEKT	1045
Qy	727	NLSY-NNVIOPGOSTTFGFGNGSYSGTNTAPTUTCT	760
Db	1046	--AYIDGLVWGOEP-----SGATPAPTPTPT	1070

```

      15.3%   Score 632.5;   DB 2;   Length 1742;
      22.8%   Post local similarity   Pred. No. 5e-24;
      143;   Mismatches 290;   Indels 335;   Gaps 37;
      7   PARAFVATAAGTAAATAATLCSITWPSA-----TAPAGFVTSGGGOFV 50
      170  PCSAVVAETAAASLAASATLVLDKRNPTAAATYVLOHAKDLFEAVYTKDSGGYTAANG----- 225
      51  LNLGPLYRGYGTNNYYLSYQSHADV-----DDVLAKAAQAMNLSVITGWGFD 96

```

RESULT 5
T31085
xylanase - Caldicellulosiruptor sp.
C:Species: Caldicellulosiruptor sp.
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31085
R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
Submitted to the EMBL Data Library, December 1997
A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: D86153
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.K.; Creasy, T.H.; Dewar, K.;
ansen, N.E.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
Rizzo, M.; Rooney, T.; Kowley, D.; Sakano, H.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
ker, M.; Salzb, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D86153

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-411 <S>O>

A:Cross-references: GB:AE005172; NID:99857528; PIDN:AAC00883.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 12.8%; Score 526.5; DB 2; Length 411;
Best Local Similarity 32.0%; Pred. No. 1.9e-19;
Matches 133; Conservative 70; Mismatches 142; Indels 71; Gaps 14;

Qy 36 AAPAGFVTASGGQFVLNGLPYRGCTNNYL-----SYQSHADVDVLAKAQAAMNLSV 88

Db 24 AVKTGFVGRNCTQFVLNCEQVYLNGFNAYNMMTTAAADTSKGRATVITATLQAASAVGMNV 83

Qy 89 IRTWGFIDIGSLDGSVPTIDGNKNGFFQYWDPTGAP-AYNDGPTGLQGLDYAIAASAAA 147

Db 84 ARINGF-----NEGDIY-----PLQISPGSYSEDV--FKGLDFVYVEAGR 121

Qy 148 HGLRVIVLTLNDKKEFGMDQYDKWYGLPYHDNFYTDPTQOAYKKNVHLLNRVNSITG 207

Db 122 FNKLIILSVNNEEDYGRKKYVEMAGLDEPDEFTYSKQFKNHVKVTLTKNTIT 181

Qy 208 VTKNDPTIFAWELANEPKRCVSGTGLPTSGCTQATIV-NVVDQMSAYVKSIDPNHMSV 266

Db 182 RMYKDDPTIFSWELINPRCND-----TASNILQDVKWKEMASVYKSIDSNHLEI 232

Qy 267 GDEGFYIGSTOGSCWPYN-----DPSDGVNALLRVKNIIDFTYHLYPNYW----- 313

Db 233 GLEGFY-GESIPERTVYVNGRVLGTGDFITNN--QIPDIDFATHIYPDSWLPLOSSR 288

Qy 314 -GNQADMGCTQWIKDIANA-AAIGKPTILEEF-----GWQTPDRDSVYQTQTQVTRN 364

Db 289 TGEODTVDWIGAHIEDCDNIKKPLITEFGKSSKYPGFSLEKRNKFFQRYVDVIYS 348

Qy 365 GEAGWN-----FWMLAGNVNQPNYDGNVYVPSSTATVLAASEALAISTGTSPP 415

Db 349 ARAGGSGCTGCVFQWLTNRTCLLDGDEVEFMQAGPNITTAQLIADOSSKLNKLYPP 404

RESULT 8

T48214

endo-1,4-beta-mannosidase-like protein - Arabidopsis thaliana

N:Alternate names: protein T20L15.200

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48214

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mey

submitted to the Protein Sequence Database, March 2000

A:Reference number: 224488

A:Accession: T48214

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-448 <BEV>

A:Cross-references: EMBL:AL162351

A:Experimental source: cultivar Columbia; BAC clone T20L15

C:Genetics:

A:Map position: 5

A:Introns: 34/2; 126/3; 191/3; 303/2

A:Note: T20L15.200

Query Match 12.1%; Score 498; DB 2; Length 448;

Best Local Similarity 32.2%; Pred. No. 5.4e-18;

Matches 132; Conservative 66; Mismatches 130; Indels 82; Gaps 17;

Qy 42 VTASGGQFVLNGLPYRGCTNNYL-----SYQSHADVDVLAKAQAAMNLSVIRTWGTFID 96

Db 48 VORKGMOFTLNGQPFYVNGFNMTTAAADNSTRGKVTVEFQOASAVGTMVGRTWAFND 107

Qy 97 IGSLOGSVPTIDGNKNGFFQYWDPTGAPAYNDGPTGLQGLDYAIAASAAHGLRVIVL 156

Db 108 -----QWRALOKSPSVYDEEV-FKALDFVLSSEARKYKIRLILSL 146

Qy 157 TNDKKEFGMDQYDKW-----YGLPYHDNFYTDPTQOAYKKNVHLLNRVNSITGVITYK 211

Db 147 VNNDAVGGKAOYVKNASGLNTSDDDFTNPTLRFYQSHVTVLUNKVNTFNITTYK 206

Qy 212 NDPTIFAWELANEPKRCVSGTGLPTSGCTQATIVNVVDQMSAYVKSIDPNHMSVYDGEF 271

Db 207 NDPTIFAWELANEPKRCPSD-----PSGDKLQ-----SWIOEMAVFKSLDAKHLVEICLEGF 258

Qy 272 YIGS-----TOGSGWPY--NDPSDGVNALLRVKNIIDFTYHLYPNYWGQNA-----DW 319

Db 259 YGSPAPARTRENPNYAAQVGTDFIRNQVL---GIDFASVIRVYVDSWISPAVSNFLEF 315

Qy 320 GTOWIKDIANA--IGKPTILEEF-----GWQTPDRD-----SVYOTWTQTVRTNGEA 367

Db 316 TSSWQAHVEDAEYGLMPVLFTEFGVSAHPDGPNTSFRMMLNTVYKMTLNSTRKGG-A 374

Qy 368 GWNFWMLAGNVNGQPNY-----YDGFNVY-----YPSSTATVLAASEALAI 408

Db 375 G-----AGSLVQVFOGAEFMDGDAYVILTRAHTASKIISLOSKRLAI 418

RESULT 9

T09048

probable mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Arabidopsis thaliana

N:Alternate names: (1-4)-beta-mannan endohydrolase; protein F26K10.200

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban

submitted to the Protein Sequence Database, June 1999

A:Reference number: 216533

A:Accession: T09048

A:Molecule type: DNA

A:Residues: 1-431 <BEV>

A:Cross-references: EMBL:AL049803; ATSP:F26K10.200; GSPDB:GN00062

A:Experimental source: cultivar Columbia; BAC clone F26K10

C:Genetics:

A:Gene: ATSP:F26K10.200

A:Map position: 4

A:Introns: 120/3; 186/3; 227/3; 298/2

C:Keywords: glycosidase; hydrolase

Query Match 12.0%; Score 495; DB 2; Length 431;

Best Local Similarity 32.9%; Pred. No. 7.3e-18;

Matches 124; Conservative 55; Mismatches 124; Indels 74; Gaps 16;

Qy 37 APAGFVTASGGQFVLNGLPYRGCTNNYLSYQSHAD-----VDDVLAKAQAAMNLSVI 89

Db 37 AKLGFVVRNGTQFVDDKPLVYVGNWSYW--FMDHAYDEHSRNLVGEHLEACAKMGLTVC 94

Qy 90 RTWGFIDIGSLDGSVPTIDGNKNGFFQYWDPTGAPAYNDGPTGLQGLDYATASAAAHG 149

Db 95 RTWAF-----NDGCYNALQI-----SPGRFDERV-FOALDHVIAEARKHD 133

Qy 150 LRVIVLTLNDKKEFGMDQYDKW-----YGL-PYHDNFYTDPTQOAYKKNVHLLNRVN 203

Db 134 VRLLSLVNNLQAYGGKTYKVAWQEGVGLSSNDSDFFDPDSIRNYFKNYLKVLLTRKN 193

Qy 204 SITGVTYKNDPTIFAWELANEPKRCVSGTGLPTSGCTQATIVNVVDQMSAYVKSIDPNHM 263

Db 194 SVTGIEYRNDPTIFAWELINEPRC-----TTDVSQ-----KTLDWDIDENTGFIKSIDOKHL 245
Qy 264 VSGDEGFYIGSTGCGS-----WPNYNDPSGDVNDNALLVKVKNIDFGTYHLYPNYWGON 316
Db 246 LTVGLEGFY-GPNSPKGLTVNPEQWASQLGTDFVQNS---NSSNIDFASVHIYPDIHWFHN 301
Qy 317 ADMG-----TOWIKDKHIANA-AAIGKPTILEERG-----WOTPDORSVYQ-----TW 357
Db 302 QTFFEEKLKFVVKWQSHIEDGLKELKRPVLTFEGLSNQNKDYEPSQDKFYRIIFDVVY 361
Qy 358 TQTVRTNGEAGNFWML 374
Db 362 KSAKKRKSAGCTLVWLQ 378

RESULT 10

S36859

c1pA protein - Clostridium thermocellum

N:Alternate names: probable cellulosome protein large chain SL

C:Species: Clostridium thermocellum

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000

C:Accession: S36859; S33527; S25767; S28659; T18261

R:Gargross, U.T.; Demain, A.L.

Submitted to the EMBL Data Library, January 1993

A:Reference number: S36859

A:Accession: S36859

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1854 <GER>

R:Gargross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.

Mol. Microbiol. 8, 325-334, 1993

A:Title: Sequencing of a Clostridium thermocellum gene (c1pA) encoding the cellulosomal

A:Reference number: S33527; MUID:93302508; PMID:8316083

A:Accession: S33527

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1615,1617-1854 <GE2>

A:Cross-references: EMBL:L08665

R:Fujino, T.; Beguin, P.; Aubert, J.P.

FEMS Microbiol. Lett. 94, 165-170, 1992

A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that b

A:Reference number: S25767

A:Accession: S25767

A:Molecule type: DNA

A:Residues: 1216-1232,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1854 <FUJ>

A:Cross-references: EMBL:X67406

R:Bequin, P.

submitted to the EMBL Data Library, August 1992

A:Reference number: S28659

A:Accession: S28659

A:Molecule type: DNA

A:Residues: 1216-1232,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1768,'R',1770-1854

A:Cross-references: EMBL:X67406

R:Fujino, T.; Beguin, P.; Aubert, J.P.

J. Bacteriol. 175, 1891-1899, 1993

A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose

e.

A:Reference number: Z18847; MUID:93209931; PMID:8458832

A:Accession: T18261

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1821-1854 <FU2>

A:Cross-references: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAA47840.1

C:Genetics:

A:Gene: c1pA

Query Match

Best Local Similarity 25.1%; Score 484; DB 2: Length 1854;

Matches 165; Conservative 102; Mismatches 208; Indels 182; Gaps 24;

Qy 229 GSGTLPTSGTCTQNTIYVWQNSA-----YKSIDPNHMSVSGDE-----GFYIGSTQG 278

Db 114 GRGTATIDGCVFATIVATVKSAAAAAPITLLEVCAFANDDLVEISTTFVACGVNLCSSVP 173
Qy 279 SCWPNYNDPSGDVNDNALLRVKNI--DFCTYHLYPNYWGONADMGQWQIWKHIANAAAI CK 336
Db 174 TTQP--NVPSCDV---VVEICKVTGCVGCTTVEIPVYFRGVPSKG-----IANDDFVFR 221
Qy 337 --PTILEEFG-----WQT---PDR-----DSVYQT 356
Db 222 YDPNVLEITIGIDPGDIIVDPNPTKSFDTAIYPRDKIIVLFAEDSGCTGAYAITKDVFAK 281
Qy 357 WTQTVRTNGEAGNFWMLAGNVNGOPYNYDGFNVVYPSSTATVLAESALAIETCT---S 413
Db 282 IRATVKSSA-----PGYITFDEVGGFAONDLEQKVSFIDKGVNWN 323
Qy 414 PPPSPSSSPSSP 473
Db 324 ATPTKGATPTNTATPTKSAATATPTRPSVPTNTPTNTANTPVSGNLKVEFYNSNPSDITN 383
Qy 474 QIKPCLQLVNTGSSSVDLSTVTYVWFTTRDGGSSSTLVVYNCDAAM-----GCGNJK 524
Db 384 SINPQFKVNTGSSAIDLSKLTLYRYITVDGOKDTFW-CDHAAIIGNSGNYNGITSNVK 442
Qy 525 ASFGSVNPTADTYLQLSFTGTLAAGSTGTEIQNRVKNKSDNSNFDDETNDYNYGTNTA 584
Db 443 GTFVKMSSTNADTYLEISFTGCTLEPCAIV-OIQGRFAKNDWSNYTOSNDYSFKSASQ 501
Qy 585 FQDWTKVTVVYVNGRLVWCTEPSGTSPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 644
Db 502 FVEWDQVTAYLNGVLVWGKEPG---SVVPSTQPVTTT-PATTKPPATTKPATTPPSD 557
Qy 645 TP-----SPSPSPSPSPSPSSGV---GCRATY-----VV 671
Db 558 DPNAIKTKVDTVNAKPGDTWIPVRFSGIPSKGIANCDFVSYSDPNVLEIIEIKPCELI 617
Qy 672 NSDMGSGF-----TATVTVTNTGSRATSGTWVASFEGGNOTVTNY 711
Db 618 DPNDKSFDTAVYPDRKIIVFLFAEDSCTGAYAITKDCVFATI----- 660
Qy 712 WNTALTOSGAS-----VTATNLSYNNVLOPQOSTTTFGNGSYSGNTAPTICT 760
Db 661 --VAKVKSAPNGLSVIKFVEVGGFANNDLVE--ORTOFFDGGVNVGDTTVPPTTPT 713

RESULT 11

G69593

cellulase (EC 3.2.1.4) bgIC precursor - Bacillus subtilis

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000

C:Accession: G69593; A26114; I40353; S24239; S49103; I39803

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinc

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.: Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: G69593

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-508 <KUN>

A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13696.1; PID:e1834

A:Experimental source: strain 168

```

Db 154 AKEFFEKMSL-----YGNTPNVJ-YEIANEP-----NCD-----VNNKRDIKPY 199
Qy 255 VKSI-----DPNHMVSGDEGFYICSTQGGCWPNDFSDGVNNAHLRVKNIIDFGTYH 307
Db 193 AEEVISVIRKNDPDNIIV-----GTCWSDVDNAADDO--LKDANVMY-AIJJ 238
Qy 308 LYPNYWGNADWGTQWTKDHIAANAAIGKPTILLEFGWQTPD---RDSVY---QTWTOT 360
Db 239 FYAGTHG-----QFLDRKANVALSKGAPIFVTE--WGTSDASGNGGVFLDOSREMIKY 289
Qy 361 VRTNCEAGNFWMLAGNVNCP-----YPNYDGFNVYPSSTATVLALEALAI-STGTSPI 415
Db 290 LDSK-TTSWNWNNLSDKOESSALKPGASKTGWRLSDSLASCTTFVRENILGTKDSTKI 348
Qy 416 PSPSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 475
Db 349 PE-----TPSKDKPTQFNCISIVQYRAGDCGSMNSNOI 379
Qy 476 KPGQLVNTGSSVDLSTVTVRYWF-TRDGGSTLTVYNCDAWAGCCGINRASFSGVNPAT 534
Db 380 RPLQIKNNGNTVDLKDVTARYWYKAKNKGON---FDCDYAQIGCGNVTHKFWTLHKPK 436
Qy 535 PTADTYLQLSFTGCTLAAGSGTGEIONRVNKSDMSNPDETNDYSGYNTAFQDMTKVTYV 594
Db 437 OGADTYLELGFKNGLTAPGASTGNIQLRLHDDHNSYAGSDYSPFKNTFTKTKKITLY 496
Qy 595 VNGRLVWGTPTS 606
Db 497 DOGKLINGTEPN 508

RESULT 12
A26874
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-19
C:Accession: A26874; B26874
R:Robson, L.M.; Chambliss, G.H.
J. Bacteriol. 169, 2017-2025, 1987
A:Title: Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.
A:Reference number: A26874; MUID:87194581; PMID:3106328
A:Accession: A26874
A:Molecule type: DNA
A:Residues: 1-508 <ROB1>
A:Cross-references: GB:M16185; NID:g143007; PIDN:AAA22496.1; PID:g143008
A:Experimental source: strain DLG
A:Accession: B26874
A:Molecule type: protein
A:Residues: 39-53 <ROB2>
A:Experimental source: strain DLG
A:Note: The authors believe Met-1 and Met-2 may be alternate initiators
C:Comment: The low molecular weight of the mature protein suggests carbon
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glu
A:Pathway: cellulose degradation
C:Keywords: alternative initiators; extracellular protein; glycosidase; l
F:1-38/Domain: (or 2-38) signal sequence #status predicted <SLC>

Query Match 11.0%; Score 454; DB 2; Length 508;
Best Local Similarity 25.7%; Pred. No. 9.9e-16;
Matches 157; Conservative 93; Mismatches 210; Indels 150; Gaps:
Qy 22 AAATLGSITMPSATAAPAGFVTASGGQFVLNG--LPYRYG-----CTNNYILSYOSHAD 7
Db 24 AVLTMGGLLPSPAAGAKTKTPVAKNGQLSTKGTQLVNRDGRAVOLKGISHSGLQWYG--- 8
Qy 74 VDDVLAKAQAAMNLSVIR-TWCFIDTGLSDGSVPTIDGNKNGFYFYQYWDPSGTGAPAYNDGP 1
Db 81 --DFVNDK---SLKWLRRDNG---ITVFRAKMYTAGD-----GYIDNP 1
Qy 133 TGLOGLDYAIASAAHGLRVIVLWLTNDWKKEFGGMDQYDKWYGLPYHDNFYDTPDPTQAYK 1

```

Db 116 SVKNKVEAAAEKELGIYVII-----DW-----HILNDGNPNQ-NK 151
QY 193 NWVHLLNRVNSITGVTKNDPTIFAMELANEPRCVSGTLPTSGTCTQATIVNWVDS 252
Db 152 EKAKKDFEKKSSSL-----YGNTPNVI-YEIANEP-----NGD-----VNKKRDIK 190
QY 253 AYVKSI-----DPNHVSVGDEGFYIGSTQSGWMPYNDPSDGVNALLRVKNIDFGT 305
Db 191 PYAEVIVSVIRKNDPDIILIV-----CTGTWSQDVNDAAADQ--LKDANVMY-A 236
QY 306 YHLYPNWGNADWGTOWIKDHIANAAGCKPTILEEFGHQTDPDRDSYVOTWTQVTRNG 365
Db 237 LIHFYAGTHGOS-----LRDKANYALSKGAPIFVTE--WGTS--ASG 274
QY 366 EAG-----WNFMLAGVNGOPYPNYDGFNVYPSSTATVLAASEALALSTCTSPPPS 417
Db 275 NGCVFLDOSREWLNLYLDSKNIS---WVWNLSLSDQKQESSALKPCASK-----TGCWPLTD 326
QY 418 PSSSPSSPSSPSSPASASPSASPSASSPSSPSSPVSGVKVQYKKNDSAPGDNOIKP 477
Db 327 LTASGTFVRENIRGKTUSTRKDPETPAQDNPTQEK-----GVSVOYKAGDGRVNSNOIRP 381
QY 478 GLOLVNTGSSVDLSTVTVRYWF--TRDGGSTLVYNCDMAAGCGNIRASFGSVNPATPT 536
Db 382 QLIHKNGNATVDLKDVTARTWYNKNGON---FDCDYAOMCGCNLTHKFTVTLHKPKQG 438
QY 537 ADTYLQLSFTGGTLAAGSGTCEIQNRVKNKSDWSNFDENDSYGNTAFQDMTKVTVVN 596
Db 439 ADTYLELGFKTGTLPCASTGNIOLRHNDWSYAGSDYSFFQSNFTKTKITLHYQ 498
QY 597 GRLVWGTPTS 606
Db 499 GKLWGTPTN 508

RESULT 13
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IF03034)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: A27198
R:Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A:Reference number: A27198; MUID:87190397; PMID:3106035
A:Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NAK>
A:Cross-references: GB:M28332; NID:gl42670; PIDN:AAA22307.1; PID:gl42671
A:Experimental source: strain IF03034
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 11.0%; Score 451.5; DB 2; Length 499;
Best Local Similarity 25.3%; Pred. No. 1.3e-15;
Matches 155; Conservative 99; Mismatches 211; Indels 147; Gaps 27;

QY 12 VATAAGTAVAAATLCSITHPSATAPAGVFTASGGQFV-LNGLPYRYGGTNNYLSYQS 70
Db 16 VLTMGGLQASPAAGTKT-----PAAKNGOLSTKGTQLVNRDGRKAVOLKGISSHGLQWYG 71
QY 71 HADVDVLAQAQAMNLSVIR-TWGFIDIGSLDGSVPTIDGNKNGFYFOYWDPTSGAPAYN 129
Db 72 -----DFVNKD---SLKWLDDNG---ITVFRAMTADG-----GYI 103
QY 130 DPTGTGLOGLDYAIAASAAHGLRIVVLTNDWKEFGMDQYDKWGLPYLHDFNTDPTQ 189
Db 104 DNPVSKNKVKEAAEAKELGIYVII---DW-----HILNDGNPNQ 140

QY 190 AYKNWVHLLNRVNSITGVTKNDPTIFAMELANEPRCVSGTLPTSGTCTQATIVNWVD 249
Db 141 -HKEKADKDFEKKSSSL-----YGNTPNVI-YEIANEP-----NGD-----VNKKR 178
QY 250 OMSAYVKSI-----DPNHVSVGDEGFYIGSTQSGWMPYNDPSDGVNALLRVKNID 302
Db 179 DIKPYABEIVSVIRKNDPDIILIV-----CTGTWSQDVNDAAADQ--LKDANVM 225
QY 303 FGYTHLYPNWGNADWGTOWIKDHIANAAGCKPTILEEFGHQTDPDRDSYVOTWTQVTR 362
Db 226 Y-ALHFYAGTHGOS-----LRDKANYALSKGAPIFVTE--WGTS-- 262
QY 363 TNGEAG-----WNFMLAGVNGOPYPNYDGFNVYPSSTATVLAASEALALSTGTS 414
Db 263 ASNGCVFLDOSREWLNLYLDSKNIS---WVWNLSLSDQKQESSALKPCASK-----TCGWP 314
QY 415 PPSPPSSPSSPSSPASASPSASPSASSPSSPSSPVSGVKVQYKKNDSAPGDNO 474
Db 315 LTDLTASGTFVRENIRGKTUSTRKDPETPAQDNPTQEK-----GVSVOYKAGDGRVNSNO 369
QY 475 IKPGLQLVNTGSSVDLSTVTVRYWF--TRDGGSTLVYNCDMAAGCGNIRASFGSVNPA 533
Db 370 IRPOLHTKNGNATVDLKDVTARTWYNKNGON---FDCDYAOMCGCNLTHKFTVTLHKP 426
QY 534 TPTADTYLQLSFTGGTLAAGSGTCEIQNRVKNKSDWSNFDENDSYGNTAFQDMTKVTV 593
Db 427 KQAGADTYLELGFKTGTLPCASTGNIOLRHNDWSYAGSDYSFFQSNFTKTKITL 486
QY 594 YVNGRLVWGTPT 605
Db 487 YHOGKLIWGTPT 498

RESULT 14
S54744
cellulase (EC 3.2.1.4) CelV1 precursor - Erwinia carotovora (SCC 3193)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Erwinia carotovora
A:Variety: SCC 3193
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S54744; S4996
R:Maee, A.; Heikkinheimo, R.; Palva, E.T.
Mol. Gen. Genet. 247, 17-26, 1995
A:Title: Structure and regulation of the Erwinia carotovora subspecies carotovora SCC
A:Reference number: S54744; MUID:95231512; PMID:7715600
A:Accession: S54744
A:Molecule type: DNA
A:Residues: 1-504 <MAE>
A:Cross-references: EMBL:X79241; NID:g493492; PIDN:CAA55823.1; PID:g493493
C:Genetics:
A:Gene: celV1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-504/Product: cellulase #status predicted <MAT>

Query Match 10.9%; Score 451; DB 2; Length 504;
Best Local Similarity 29.0%; Pred. No. 1.4e-15;
Matches 142; Conservative 64; Mismatches 158; Indels 126; Gaps 24;

QY 143 ASAAHGLRIVVLTNDWKEFGMDQYDKWGLPYLHDFNTDPTQOAYKNWVHLLNRV 202
Db 114 AVAAAGLGVYIIII--DWHTLS-----DN-----YKAQAKIFFAEM 151
QY 203 NSITGVTKNDPTIFAMELANEPRCVSGTLPTSGTCTQATIVNWVDSAY-----V 255
Db 152 AGL-----YGNSPNVI-YEIANEP--GS-----VTWNCQIRPYALEVTDTI 190
QY 256 KSTDHPNHVSVGDEGFYIGSTQSGWMPYNDPSDGVN-----NALLRVKNIDFGTYLHP 310
Db 191 RSKDPONLIIV-----GSGTWSQDIHDAADNQLDPNLTVAL-----HIFY 231

[illegible]

Search completed: November 13, 2002, 11:57:31
Job time : 41.5819 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:47:41 : Search time 17.5172 seconds
(without alignments)
1804.220 Million cell updates/sec

Title: US-09-917-378-1

Perfect score: 4122

Sequence: 1 MGLVRRPARAFVATAAGTAV.....GFNGSYSGTNTAPTTLCTAS 762

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696.5	16.9	562	1 GUN1_ACICE	P54583 acidothermu
2	676.5	16.4	1331	1 MANB_CALSA	P22533 caldocellum
3	632.5	15.3	1742	1 GUNB_CALSA	P22534 caldocellum
4	603.5	14.6	1039	1 GUNB_CALSA	P10474 c endogluc
5	484	11.7	1853	1 CIPA_CLOTH	Q06851 clostridium
6	476	11.5	772	1 CIPB_CLOTH	Q01866 clostridium
7	460	11.2	499	1 GUN2_BACSU	P10475 bacillus su
8	454	11.0	499	1 GUN1_BACSU	P07983 bacillus su
9	451	10.9	504	1 GUNW_ERWCA	Q59395 erwini
10	451	10.9	747	1 GUNW_ERWCA	P50400 cellulomona
11	447	10.8	499	1 GUN3_BACSU	P23549 bacillus su
12	445	10.8	700	1 GUNA_PAELE	P29719 paenibacill
13	426.5	10.3	505	1 GUNV_ERWCA	Q47096 erwini
14	394	9.6	145	1 YCEA_PAELE	P29718 paenibacill
15	380.5	9.2	456	1 GUNA_MICBI	P26414 microbisp
16	380	9.2	441	1 GUN2_THEFU	P26222 thermomonos
17	376.5	9.1	880	1 GUN4_THEFU	P10476 pseudomonas
18	362	8.8	962	1 GUNA_PSEFL	P26225 cellulomona
19	335	8.1	1045	1 GUNB_CELFI	P50401 cellulomona
20	330	8.0	872	1 GUXA_CELFI	P50900 clostridium
21	315	7.6	914	1 GUX2_CLOS	P50899 cellulomona
22	313.5	7.6	1090	1 GUXB_CELFI	P23659 cellulomona
23	311.5	7.6	484	1 GUX_CELFI	P23659 cellulomona
24	308.5	7.5	986	1 GUNZ_CLOS	P01034 orgyia pseu
25	300	7.3	279	1 Y091_NPVOP	Q02934 clostridium
26	298	7.2	879	1 GUN1_CLOTH	P19487 xanthomonas
27	287	7.0	484	1 GUNA_XANCP	P19487 xanthomonas
28	285.5	6.9	474	1 VTP3_TTVIV	P19487 xanthomonas
29	274	6.6	555	1 GPI_CHURE	P19487 xanthomonas
30	268	6.5	515	1 GUN2_CLOCL	Q9fpg6 chlamydomon
31	252.5	6.1	1229	1 N121_HUMAN	Q942n3 homo sapien
32	245	5.9	1367	1 AMYH_YEAST	P08640 saccharomyc
33	245	5.9	1848	1 CBPA_CLOCL	P38058 clostridium

ALIGNMENTS

RESULT 1

GUN1_ACICE	STANDARD;	PRT;	562 AA.
ID	GUN1_ACICE	STANDARD;	PRT;
AC	P54583;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)		
DE	(Cellulase E1) (Endocellulase E1).		
OS	Acidothermus cellulolyticus.		
OC	Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;		
OC	Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.		
OX	NCBI_TaxID=28049;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 43068 / 11b;		
RA	Laymon R.A., Himmel M.E., Thomas S.R.;		
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.		
RX	MEDLINE=96346058; PubMed=8718854;		
RA	Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplus P.A.;		
RT	Crystal structure of thermostable family 5 endocellulase E1 from		
RT	Acidothermus cellulolyticus in complex with cellobiose.		
RL	Biochemistry 35:10648-10660(1996).		
CC	- - FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81		
CC	DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON		
CC	CARBOXYMETHYLCELLULOSE.		
CC	- - CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic		
CC	linkages in cellulose.		
CC	- - SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL		
CC	HYDROLASES).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL: U33212; AAA75477.1; -		
DR	PDB; LECE; 14-OCT-96.		
DR	InterPro: IPR001919; Bac_cellose-bind.		
DR	InterPro: IPR001547; GH_5.		
DR	Pfam: PF00150; cellulase; 1.		
DR	Pfam: PF00553; CBM_2; 1.		
DR	PROSITE: P500659; GLYCOSYL_HYDROL_F5; 1.		
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.		
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.		
FT	SIGNAL 1 41		
FT	CHAIN 42 562		
FT	DOMAIN 42 400		
FT	CATALYTIC		
FT	PRO/SER/THR-RICH (LINKER).		
FT	DOMAIN 401 461		
FT	CELLULOSE-BINDING (BY SIMILARITY).		
FT	DOMAIN 462 562		
FT	ACT_SITE 203 203		
FT	ACT_SITE 323 323		
FT	NUCLEOPHILE.		

Q59394 erwini car
P22698 dictyosteli
P27033 pseudomonas
P16356 raenorhabdi
Q02817 homo sapien
P36909 streptomyce
P54865 cellulomona
P40602 arabidopsia
P07984 cellulomona
Q06852 clostridium
P11220 streptomyce
P74745 synchocyst.


```
Query Match 16.4%; Score 676.5; DB 1; Length 1331;
Best Local Similarity 24.5%; Pred. No. 2.8e-25;
Matches 208; Conservative 135; Mismatches 262; Indels 243; Gaps 33;

OY 60 GTNNYLSYQSHADVDVLAKAQAAMLSVITWGTFTDGLSLDGSVPTIDGNKNGFYQYW 119
DB 55 GTNHACWYRDLDT-----ALRGIRSGM-----NSRVV-----LSNGY---RW 92

OY 120 DPTGAPAYNDGPTGLOGLDYAATASAAHCLRVIVVLND-----W 160
DB 93 TKIPASEVAN-----IISLSRLGFKAILIEVHDTTCYGEDGAACSLAQAVEYW 141

OY 161 KEFGG-MDYQDKWGLPYLHDFNYDPTQOAYKNWYHLLNRVNSITGVYKN---DPT 215
DB 142 KEIKSVLDGNEDEVII---NIGNEPYGNNNYQWYNDTKNAIKALRDAGFKHTIMVDAP 197

OY 216 IFAMELANEPRCVSGTGLTSG-----TCTQATVNVVDQMSAYVKS-IDPNHMYSVGD 268
DB 198 NMGQDSNTRMDNAQAIMEADPLNLVFSITHMYGVYNTASKVEEYIKSEVDKGLPLVIGE 257

OY 269 EGFYIGSTGSGHPYNDPSDGVNNALLRVK-----NIDECTYHLYPNYWC 314
DB 258 FGHQ--HTDG-----DP-----DEAIVRYAKOYKICLFSWSCGNSVYGVLDWVNNMD 305

OY 315 QN--ADWGTQWIKDHIANAAGIGKPT-----PNYDC 388
DB 306 PNNPTPWG-QWYKTNALGTSSTPTPTSTVTPPTPTPTPTPTPTPTPTPTPTPTPTPTPT 364

OY 339 -----ILEERG-----WOTPRD---SVYQWTQ 359
DB 365 GQIKVLYANKETNSTTNTIRPWLKVNVSGSSSIDLSRVTRYTYTVGDGERAQSAISDWAQ 424

OY 360 TVRTNCEAGWFMWLAGNVGOPY-----PNDYD 388
DB 425 IGASN-VTFKFKVLSVSGGADYILEIGFKSAGQLOPKQKOTGEIQMRNKDDWSNYNQ 482

OY 389 FN-----VYPSSTATVLAELAIS---TGTSPSPSPSSPSPSPSPSPSPSPSP 436
DB 483 GNDWSIQSMTSYGENEKTAYIDGVLVNGQEPGSGATPAPATPTPTPTPTPTPTPTPTPT 542

OY 437 ----SASPASSPSPSPSPSPSPSGG-VKQYKNNDSAGDNOIKPGLQVLNTGSSVDL 491
DB 543 TPTVTATPTPTPTPTPTPTPTATGQIKVLYANKETNSTNTIRPWLKVNVSGSSSIDL 602

OY 492 STVTRYWFTTRDGGSTLVYNCDAAMCGCNIRASFGSVNPAATPTADTYLQLSFTCGT-- 549
DB 603 SRVTIRYWTVDGERAQSAIS-DWAIGASNVTFKEVKLSVSSVSGADYILEIGFKSAGQ 661

OY 550 LAAGSTGEIQNRVKNKSDNSDETDNDYSGTNTAFQDMTKVTVYVNGRLVNGTEPSPGT 608
DB 662 LOPGKDTGEIQIRFNKSDNSNYNOGNDWSIQSMTSYGENEKTAVIDGVLVNGQEPSPGT 721

OY 609 SPSPSPSP---SPTPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 664
DB 722 TSPSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 781

OY 665 CRATYVNVSDWSGEFTATVTV-TNTGSRATSGWTVAMSGFQNOTVTNYNWTALTQSCASV 723
DB 782 SPSPSVVE-----ITINTNAGRTQIS-----PYIYGANQDIECVVHSARRLGCNRL 827

OY 724 TATNLSYN 731
DB 828 TGYNNWENN 835

RESULT 3
GUNA_CALSA
ID GUNA_CALSA STANDARD; PRT: 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
```

```
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'O V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -!- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L32742; AAA91086.1; -
DR EMBL: M36063; AAA72860.1; -
DR EMBL: L01257; -; NOT_ANNOTATED_CDS.
DR PIR: A43745; A43745.
DR HSP: P26221; ITF4.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR000556; Glyco_hydro_48.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR Pfam: PF00942; CBM_3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR PRINTS: PR00844; GLHYDRLASE48.
DR ProDom: PD001947; CBD_3; 2.
DR ProDom: PD011903; Glyco_hydro_48; 1.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00598; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1050 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
FT SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;
```


Query Match	15.3%	Score 632.5	DB 1	Length 1742
Best Local Similarity	22.8%	Pred. No. 4.6e-23		
Matches 227	Conservative 143	Mismatches 290	Indels 335	Gaps
Qy	7	PARAFVTAAGTAAVAAAATLGSITMPSA-----TAAPAGFVTASGQGV	50	
Db	170	PCSAAVETAAASLAASIVLKDORPHTKAATYLOHAKDLFEFAEVYTKSDSGYTAANG-----	225	
Qy	51	LNGLPYRYGGTNNYYLSYQSIADV-----DDVLAKAQAMNLSVIRTWGFI	96	
Db	226	-----YVNSWGFYDELSWAAVWLYLATNDSTYLTKAE-----SYVQNW-----	264	
Qy	97	IGSLDGSVPTIDGNK-----NG-----FFYQYWDWST	123	
Db	265	-----PKISGNIIDYKWAHCWDDVHNGAALLAKITDKDTYKQIIESHLDYWTGY	316	
Qy	124	CAPAYNDGPTGLOGLD-----YATASAAHGLRVIVLTDWKFEGGDDQYDKWYGLPY	177	
Db	317	NGERIKYTPKGLAWLDQWGLRYATTATP--LAFV-----YDSMSGCP--	357	
Qy	178	HDNFYDPTQOAYKKNWVNHLLNRVNSITGVVYKNDPTIFAWELANPEPCVGSGLTPTSG	237	
Db	358	-----TGKKETRYKFGEQIDYALGSTCRSP-----VVGFGTNPPKR	394	
Qy	238	TCTQATVNVWQDMSAVKSIDPNHNVSVGDEGFYIGSTOGSGWPYNPDSGDVDNALLR	297	
Db	395	PHHRTAHSWADS-----QSIPSYHRHIL--YCALVGG--PGSDDSYTDIISVYVNEVAC	446	
Qy	298	VKNIDF-----GTVHLY-----PNYWCQADMGTOWKIDHIANAAIG-----KPTILEEF	343	
Db	447	DYNAGFYCALAKMYLLYGGNPIPDFKAIETPTNDEFVFEAGINASCTNFIETKAIANNQS	506	
Qy	344	GW-----OTPDSDSVQYOTWTQVTFNCEAGNW-----FWMLA	375	
Db	507	GWPARATNKLKRYFYVDLSELIKAGYSPNQLTLSTNYNOCAKVPYVWDSRNTIYILV	566	
Qy	376	GNVNGQYP-----NYDGFNVY-----YPSSTATVILASEALAI--	408	
Db	567	DFTCTLIYPGQDKYKKEVGFRIAAPQNVQWDSNDYSFODIKGVSSGSVVKTKYIPLYD	626	
Qy	409	-----STCTSP-----PPSPSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS	451	
Db	627	EDIKWGEERGCTSGVSPPTASVTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT	686	
Qy	452	SSP-----VSGVKVQYKKNDSAPCDNQIKPGLQVMTGSSVDLSTVTVRY	498	
Db	687	STPTVTPPTPVSTPATSCQIKVLYANKETNSTTNTLRPLKVVNSGSSIDLSRVTIRY	746	
Qy	499	WFTPDGSSLLVYNCWDAAMCCNIRASFGSVNPATPTADTYLQLSFTGGT--LAAGGST	556	
Db	747	WYTVDGERAQSALS-DWAQIGASNVYTFKVKLSSSVSGADYYLEIGFKSGAGQLQPKGDT	805	
Qy	557	GEIQNRVKNKSDWNFDETNDYSY-GTNTAFQDWTKVTVVNGRLVMGTEPESGTSPTSPPTPS	615	
Db	806	GEIQIRNKDDNSNYNOGNDWSHQSWTSYGENEKTATYIDCVLVWGOEPTGTTPTAPTPT	865	
Qy	616	P-----SPTSPSPSPPTSPSSP	652	
Db	866	PTVTVTPPT	925	
Qy	653	-----SPSPSVSSSGVGCRTAVVN-----SDWSGGFTATVT-----VNTFGS	690	
Db	926	IRPLKVVNSGSSIDLSRVTIRYVTVYDGERAQSALSDWAQIGASNVYTFKVKLSSSVS	985	
Qy	691	RATSGMTVAWSFGCQNTV-----TNY-----WNITALTOSGASVTAT	726	
Db	986	GADYYLEIGFKSGAGQLQPKGDTGEIQIRFNKDDNSNYNOGNDWSHQSWTSYGENEKT	1045	
Qy	727	NLSY-NNVQPGOSTTFFGNGSYSGTNTAPTTLCT	760	
Db	1046	--AYIDCVLVWGOEP-----SCATPAPTPTPT	1070	

RESULT 4

IND_GUNB_CALSA	STANDARD;	PRT;	1039 AA.
AC	P10474;		
DT	01-JUL-1989 (Rel. 11, Created)		
DT	01-JUL-1989 (Rel. 11, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Endoglucanase/exoglucanase B precursor [Includes: Endoglucanase (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase) (Cellobiohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)]		
DE	(1,4-beta-cellobiohydrolase)]		
GN	CELUB.		
OS	Caldocellum saccharolyticum (Caldocellulosiraptor saccharolyticus).		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;		
OC	Caldocellulosiraptor.		
OX	NCBI_TaxID=44001;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=89098398; PubMed=2789517;		
RA	Saul D.J., Williams L.C., Love D.R., Chamley I.W., Berquist P.I.;		
RT	"Nucleotide sequence of a gene from Caldocellum saccharolyticum		
RT	encoding for exocellulase and endocellulase activity.";		
RL	Nucleic Acids Res. 17:439-439(1989).		
CC	-1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL		
CC	DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS		
CC	AN ENDOGLUCANASE.		
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic		
CC	linkages in cellulose.		
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages		
CC	in cellulose and celotetraose, releasing cellobiose from the non-		
CC	reducing ends of the chains		
CC	-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY		
CC	F (FAMILY 10 OF GLYCOSYL HYDROLASES).		
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY		
CC	A (FAMILY 5 OF GLYCOSYL HYDROLASES).		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EWBL: X13602; CAA31936.1; ..		
DR	PIR: S02711; S02711.		
DR	HSSP: O06851; INBC.		
DR	InterPro: IPR001956; CRD_3.		
DR	InterPro: IPR001547; GH_5.		
DR	InterPro: IPR001000; Glyco_hydro_10.		
DR	Pfam: PF001150; cellulase; 1; 10.		
DR	Pfam: PF00331; Glyco_hydro_10; 1.		
DR	Pfam: PF00942; CBM_3; 1.		
DR	PRINTS: PR00134; GLYHYDRLASE10.		
DR	ProDom: PD001947; CBD_3; 1.		
DR	PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.		
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.		
KW	Cellulose degradation; Hydrolase; Glycosidase; Repeat;		
KW	Multifunctional enzyme; Signal.		
FT	SIGNAL 1 28		
FT	CHAIN 29 1039		
FT	DOMAIN 376 416		
FT	DOMAIN 417 570		
FT	DOMAIN 571 618		
FT	ACT_SITE 177 177		
FT	ACT_SITE 285 285		
FT	ACT_SITE 792 792		
SQ	SEQUENCE 1039 AA: 117641 MW; OE0378171594DDAE CRC64;		

Query Match 14.6%; Score 603.5; DB 1; Length 1039;

Best Local Similarity 28.5%; Pred. No. 6.3e-22;

Matches 179; Conservative 107; Mismatches 209; Indels 133; Gaps 30;

Query Match 14.6%; Score 603.5; DB 1; Length 1039;
Best Local Similarity 28.5%; Pred. NO. 6.3e-22;
Matches 179; Conservative 107; Mismatches 209; Indels 133; Gaps 30;

QY 160 WKEFGMDQYDKWYGLPYHDNEYDTPRTOQAYKNVNHLLNRVNSITGVYTKNDPTIFAW 219
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 192 WYICGPEYIEKAF-IWAHE---ADPNAKLFYNDY-----NTEISKKRD---FIY 234
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 220 ELANEPRCVCGLTPTSGTCTOATI-VNW--VDOMSAYVK--SIDPN---HMVSVGDEGF 271
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 235 NMVKNLKSKG---IPIHGICMOCHINWNPSEIENSIFKLFSSIPGIEIHTEL-DMSL 290
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 272 YIGSTOGSCWPNYNDPSGDVGNALLRVKNIDFGTYHILYPN-----YWCQNAWGTOWIK 335
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 291 Y---NYVCSSSENYTPPDQLLOKQSKYKEI-FTMLKKYKNVKSVTFTGLKDDY--SWLR 344
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 326 DHIAAAAIGKPTILIEFCGQOTDRDSVYQTWTQTVRTNCEAGNFWMLAGVNGQYPN 365
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 345 -----SFF-----CKNDWPLL----- 355
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 386 YDGFNVYPSSTATVLAEEALISTGTSPPSPSSPSSPSSPSSPSSPSSPSSPSSPSS 445
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 356 ---FFEDYSAPAYMAVIEASGVT--SSPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 407
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 446 PSPSPSSPVSGG-VKVOYKNNDASPCDNOIKPGLQLVNTGSSSVOLSTVTVRYWFTRDG 504
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 408 PTPPTVSTPATGGOIKVLYANKENTSTNTIRPWLKVYVNSGSSIDLSRVTIKYWTVDG 467
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 505 GSSTLVYNCDAAMCGCINIRASFGSVNPATPTADTVLQLSFTGCT--LAAGSGTGEIONR 562
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 468 ERAQSAYS-DWAQIGASNVTFKFKLVSSVSGADYLEIGCFKSGAGOLQPKDTGEIQIR 526
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 563 VNKSDMSNDETDNYSY-GTNTAFQDWTKVTVYVNGRLVNGTSPGTSPTSPSPSPSP 621
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 527 FNKSDMSNNOGNDMSWLOSMTSYGENEKVTAYIDGVLVNGQEPSPGATPAPTMTVAPT-- 584
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 622 PSPSPSPSSPSSPSSP 677
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 585 ATPTPLSPVTPPTAPTQTAIPTLTNPPTTSSIPDD-----TNDQWLVSXGN 635
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 678 GFT-----ATVTVT-----NTGSRATSGWTVAMSGGNTQVTNYNWTALTQSGASVTA- 725
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 636 KIVDKDGRPWLTGINKFGYNTGNTVFDG---VWSCNLKDTLAEIANRGNLRLVRPISAE 692
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 726 TNLVSNVITOPGOSTTFGNGSVSGTNT 753
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 693 LILNWSQGYPKPNIINYVNPPELEGRNS 720
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
RESULT 5
CIPA.CLOTM STANDARD; PRT: 1853 AA.
AC 006851;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (Cellulosomal
glycoprotein SL/SL) (Cellulose integrating protein A) (Cohesin).
GN CIPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
RC STRAIN-ATCC 27405 / DSM 1237;
RX MEDLINE=93302508; PubMed=8316083;
RA Geringross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
RA Demain A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cipa) encoding the
RT cellulosomal SL-protein reveals an unusual degree of internal
RT homology.";
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RP SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE=93209931; PubMed=8458832;

RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein Cipa and a protein possibly involved
RL in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
[3]
RN X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE=97238934; PubMed=9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
RA Frolow F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly.";
RL Structure 5:381-390(1997).
[4]
RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Tormo J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Steitz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose.";
RL EMBO J. 15:5739-5751(1996).
[5]
RN X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402065;
RA Tavares G.A., Beguin P., Alzari P.M.;
RT "The crystal structure of a type I cohesin domain at 1.7-A
RT resolution.";
RL J. Mol. Biol. 273:701-713(1997).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150
CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC
CC COMPONENTS OF THE CELLULOSE.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L08665; ; NOT-ANNOTATED_CDS.
DR EMBL: X67506; CAA47840.1; ;
DR PIR: S36859; S36859.
DR PDB: 1ANU; 23-JUL-97.
DR PDB: 1AOH; 08-JUL-98.
DR PDB: 1NBC; 26-SEP-97.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002102; Cohesin.
DR InterPro: IPR002105; Dockerin_1.
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00963; Cohesin; 9.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00448; CLOS_CELLULOSE-RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
KW 3D-structure.
FT SIGNAL 1 28
FT CHAIN 29 1853
FT DOMAIN 364 522
FT DOMAIN 54 1694
FT REPEAT 29 182
FT REPEAT 183 322
FT REPEAT 560 704
FT REPEAT 724 866


```
Db 428 OGADTYLELGFKNGTLAPGASTGNIQLRHNDMSNVAQSGDYSPFKSNFTKTKITLY 487
QY 595 VGNRLVWGTEPS 606
Db 488 DOCKLINGTEPN 499

RESULT 8
GUN1_BACSU STANDARD: PRT: 499 AA.
AC P07983;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (cellulase).
GN BGLC OR GLD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=DLG;
RA MEDLINE=87194581; PubMed=3106328;
RX Robson L.M., Chambliss G.H.;
RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
RL J. Bacteriol. 169:2017-2025(1987).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: M16185; AA22496.1; ALT_INIT.
DR PIR: A26874; A26874.
DR HSSP: 085465; 1A3H.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00942; CBM_3; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 237 237 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 330 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA: 55187 MW: 339004EE95A63EE1 CRC64;

Query Match
Best Local Similarity 25.7%; Pred. No. 3.7e-15;
Matches 157; Conservative 93; Mismatches 210; Indels 150; Gaps 27;

QY 22 AAATLGSITMPSATAAGFVTSAGGQFVLNG--LPYRYG-----GTNNYLSYQSHAD 73
Db 15 AVLTMGGLLPSPASAGTKTPVAKNQGLSIKGTQLVNRDQKAVQLAGISGHLQWYG--- 71
QY 74 VDDVLAKAQAAMLSVIR-TWGEDIDGSLDGSVPTIDGNKNGFYQYWDPSGTGAPYNDGP 132
Db 72 --DFVNKD---SLKWLRRDWMG---ITVFRAMRYTADG-----GVIDNP 106
QY 133 TGLQGLDYAIAASAAHGLRVIVLTNDKKEFGMDQYDKWGLPYHDNFTDPTDPTQOAYK 192
Db 107 SVKNKVKAEVAEKELGIIVII-----DW-----HILNDGNPNQ-NK 142
```

```
QY 193 NWVNHLLNRVNSITGYTYKNDPTIFAWELANEPRCVSGTLPTSGTCTOATINVNWIDMS 252
Db 143 EKAKEFFKEMSSL-----YGNTPNVI-YEIANEP-----NCD-----VNMKKDIK 181
QY 253 AYVKSI-----DPNIHMSVGDGEFYIGSTQSGWPYNDPDSGVDNALLRVKNIDFGT 305
Db 182 PYAEVIVSVIRKNDPDNIIV-----GTGTHSQDVNDAADDQ--LKDANVWY-A 227
QY 306 YHLYPNYMGONADWGTQWIKDHIANAAAICKPTILEFFGWOTPDSDSVYOTWTOTVYING 365
Db 228 LHFYAGTHGQS-----LRDKANYALSKGAPFVTE--WGTSQ-----ASPS 265
QY 366 EAG-----WNFMLAGNVNGOPNYDGFNYYVPSSTATVLALEALAISTGTSPPS 417
Db 266 NGGVFLDQSRWLNLYDSKNIS---WVWNLSDKQESSALKPKGASK-----TGGWPLITD 317
QY 418 PSSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 477
Db 318 LTASGTFVRENIRGTRKDTKDPETPAQDNPTQOEK-----GVSVOYKAGDGRVNSNOIRP 372
QY 478 GLQLVNTGSSVDLSTVTVRYWF--TRUGSGSTLYVNCDMAAMCGNIRASFGSVNPATPT 536
Db 373 QLHIKNGNATVDLKDVTARYWYNVKNKQN---FDCDYAQMCGCNLTHTKFTVLIKPKQG 429
QY 537 ADTYLQSLFTGGTLAAGSGTGEIQNRVKNKSDWSNFDENDYSYCTNTAFODWTKVTVVN 596
Db 430 ADTYLELGFKTGLTSPGASTGNIQLRHNDMSNVAQSGDYSPFSQNTFTKTKITLYHQ 489
QY 597 GRLVWGTEPS 606
Db 490 GKLIWGTEPN 499

RESULT 9
GUNW_ERWCA STANDARD: PRT: 504 AA.
AC Q59395;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
DE (Cellulase V1).
GN CELV1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinhelmo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
RT phytopathogenicity.";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SUBCELLULAR LOCATION: Secreled.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: X79241; CAA55823.1;
DR HSSP: 085465; 1A3H.
```

```
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; CH_5.
DR Pfam: PF00150; cellulase: 1.
DR Pfam: PF00942; CBM_3; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 504 ENDOGLUCANASE V1.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 504 AA: 54963 MW: 0D7ECF74781565FA CRC64;

Query Match 10.9%; Score 451; DB 1; Length 504;
Best Local Similarity 29.0%; Pred. No. 5.1e-15;
Matches 142; Conservative 64; Mismatches 158; Indels 126; Gaps 24;

QY 143 ASAAAHGLRVLTNDWKEFGMDQYDKWGLPYHDNEYTDPRTOQAYKNVNHLLNRV 202
DB 144 AVAAAGLGVYIII--DWHTLS-----DN-----DPT--YKAQAKIFFAEM 151
QY 203 NSITGVVTKNDPTIFAWELANERPCVSGTLPTSGTCTQATIVNVVDMSAY-----V 255
DB 152 AGL-----YGNSPNVI-VEIANEPN--GS-----VTWNGQIRPYALEVTDTI 190
QY 256 KSIDPNHMVSGDEGFFVIGTSGCWPNPDSDGVN-----NALLRVKNIDFGTYHLP 310
DB 191 RSKDPNDLIIV-----CGSTWSQDIHDAADNOLPDNPLYAL-----HFYA 231
QY 311 NYMGONADWCTQWIKHIANAATGKPTILEEFCQWTPDRDS-----VYQTWQTQVRT 363
DB 232 GTHG-----QFLRDIRIDVAQSRGAIFVSE--WGTSDASNGGPFLESOTWIDFLNN 282
QY 364 NGEAGNFWMLAGNVNQPNYDGFVYPSSTATVLAELAISTGTSPPPSPSSPS 423
DB 283 RG-ISMVWNSLSDK-----SETSAALVA-----GASKSGGWTEQNL 317
QY 424 SSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 479
DB 318 STSGCKFVREIQIRAGAGLSGGDTPTMTPEPTNPGNGITGDIVLQVIRVNDNPSDDAI 377
QY 480 QLVNTGSSVLDLSTVTVRYVTRDQ--GSSTLVVNCDMAAGCGNIRASFGSVNPATPTA 537
DB 378 NIKNTGSTPIKLSDLQVRYVYFHDGKPGANLV--DMANVGPNPIVSTGTTPAASDTKA 434
QY 538 DTYLQLSF-TGGTLAAGCGTGEIQNRVNSKDSWNFDETNDYSYGTN-TAFODWTKVTVV 595
DB 435 NRYVLVTFASGSLQPGAETGEVQVRIHAGDSNVNENYDYSYGNITSVTNMDKITVHD 494
QY 596 NGRVLMGTEP 605
DB 495 KGLVWGTEP 504

RESULT 10
GUND_CELFI
ID GUND_CELFI STANDARD; PRT; 747 AA.
AC P50400;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CEND.
OS Cellulomonas fimi.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococciaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
```

[illegible]

RESULT 11

```

UN31_BACSU
ID      GUN3_BACSU      STANDARD;      PRT;      499 AA.
AC      P23549;
DT      01-NOV-1991 (Rel. 20, Created)
DT      01-NOV-1991 (Rel. 20, Last sequence update)
DT      01-FEB-1995 (Rel. 31, Last annotation update)
DE      Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE      (Carboxymethyl-cellulase) (CMCase) (Cellulase).
GN      BGLC.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=1423;
[1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=BSE616;
RX      MEDLINE=91299280; PubMed=1368694;
RA      Park S.H., Klm H.K., Pack M.Y.;
RT      "Characterization and structure of the cellulase gene of Bacillus
RL      subtilis BSE616.";
RL      Agric. Biol. Chem. 55:441-448(1991).
CC      -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC      linkages in cellulose.
CC      -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC      HYDROLASES).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@lsb-sib.ch).
-----
CC      EMBL: D01057; BAA00859.1; -.
CC      PIR: JN0111; JN0111.
CC      HSP: O85465; 1A3H.
CC      InterPro: IPR001956; CBD_3.

```

DR	InterPro: IPR001547; GH5.
DR	Pfam: PF00150; Cellulase_1.
DR	Pfam: PF00942; CBM_3_1.
DR	ProDom: PD001947; CBD_3_1.
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT	SIGNAL 1 29
FT	CHAIN 30 499
FT	ACT_SITE 169 199
FT	ACT_SITE 257 257
FT	DOMAIN 350 499
FT	SEQUENCE 499 AA; 55169 MW; 2E821E3D8BACA04 CRC64;
QY	Query Match 10.88; Score 447; DB 1; Length 499;
DB	Best Local Similarity 24.4%; Pred. No. 7.9e-15;
MATCHES	Matches 149; Conservative 88; Mismatches 216; Indels 158; Caps
Qy	25 TLGSITPSTATAAPAGFVTASGGQVFLNG--LPKRYG-----GTNNYYLS-YQSIIADVD 75
Db	18 TMGGMLASPASAAGCTTPVAKNGOLSIKGTOLYNRDGCAVOLKGISIHGLQHWGEYVNKD 77
Qy	76 DVLAKAAMNLSVIR-TWGFFIDIGSLDGSVPTIDGNKNRGFYFOYWDPSFGAPAYNPGTC 134
Db	78 -----SLKLRLDWG---ITVFRAAMYTAGD-----CIIIDNPSPV 104
Qy	135 LQGIDYAIASAAAHCLRVIVLTNDWKFEFGMDQYDKWGLPYIIDNFYTDPRTQQAYKNW 194
Db	109 KNKKKEAVEAKEUGITVII----DM-----HILINDGNPNQ-NKEK 144
Qy	195 VNHLNLNRYSITGYTKNDPTIFAMELANEPRCVSGSUTLPTSCTCTOATIVNVNVDMSAY 254
Db	145 AKEFPEKMSSL-----YGNTPNVI-YETANEP-----NCGD-----VNNKRDIKPY 184
Qy	255 VKSI-----DPNHMYSVGDEGFYICSTOGSGWPYPNDPSGDYDNALLKVKNIIDKCTTH 307
Db	184 AEEVISVIRKNDDPNIIIV-----GTGTWSODVNDAAADDQ--LKDAHV-MDALII 229
Qy	308 LPNYWGNQADMGQTWIKDHIAANAATCKPTILEEFCWTQPD---RDSVV---QTWTQT 360
Db	230 FYACTHG-----QFLRDKANVALSKAPIFYTE--WCTSDASGCGVPFLDOSREMLKY 280
Qy	361 VRTNGEAGNWFMLAGNVNQP-----YPNYDGFNVIYPPSSSTATVLAEMALISTGTSP 415
Db	281 LDSC-TISWNWNLSLDKQESSALKPKGASKTGGWRLLSDLASGTFVRENILGTKDTKDI 339
Qy	416 PSPSSPSSP 475
Db	340 PE-----TPAKDKPTQENGISVOYRAGDGSMMNSNOI 370
Qy	476 KPGQLQVNTGSSVDLSTVTVRVWFTRDGSSTLVYNCDMAAGCGNIRASFQGVSNPATP 535
Db	371 RPLQIQIKNNGTTVDLKDVTARYWYNAKKNQNQV--DCDYAQLGCCGNVTVKFVTLHKPKQ 428
Qy	536 TADTYLQSLFTGGTLAGGSTGETIONRVNKSDMSNFDENITYSYGNTAFODMTKTVTVV 595
Db	429 GADTYLELGFANGTLAPCASTGNQLRLHNDWSNTAQSGDYSFFFKTSFTKTKTKITLYD 488
Qy	596 NGRLVWGTGPS 606
Db	489 QGKLINGTEPN 499
RESULT 12	
GUNA_PAELA	STANDARD; PRTr; 700 AA.
ID	GUNA_PAELA
AC	P29719;
DT	01-APR-1993 (Rel. 25, Created)
DT	01-APR-1993 (Rel. 25, Last sequence update)
DT	15-JUL-1998 (Rel. 36, Last annotation update)
DE	Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE	(Cellulase A) (EG-A).
OS	Paenibacillus lautus (Bacillus lautus).

RESULT 12

RESOL 12
GUNA PAELA

ID GUNA PAELA

AC P29719; CONTAINER STANDARD; FRI, 100 AM

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation updated)

DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-

DE (Cellulase A) (EG-A).

GN CELA.

OS Paenibacillus lautus (Bacillus lautus).

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Biderichsen B., Joergensen P.L.;
RT "Cella from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase."
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -!- PTM: A SHORT FORM (ECA-S) ARISES FROM POSTTRANSLATIONAL
PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M76588; AAA22303.1; -
DR JIR; B41897; B41897.
DR PIR; S27499; S27499.
DR HSSP; O06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 700 ENDOGLUCANASE A.
FT DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 213 213 BY SIMILARITY.
SQ SEQUENCE 700 AA; 76910 MW; 3D5C8CAD53EE0F CRC64;

Query Match 10.8%; Score 445; DB 1; Length 700;
Best Local Similarity 25.2%; Pred. No. 1.4e-14;
Matches 191; Conservative 106; Mismatches 243; Indels 218; Gaps 36;

QY 5 RRPARFVTAAGTAAVAAATLGSITMPSATAAPAGFVTASGGQFVNLGLPYRYGGTNNY 64
DB 4 ROKKRLFVS-----AALAVSLTM-TVPMPASVNAASDV-----FTINTOSERAAISPNI 53
QY 65 YLSYQSHADVDVLAQAQNN-----LSVIRTWGFIDIGSL--DGSVPTTIDGNK 111
DB 54 YGTNODLSCTENSSRRKCGNRLTGYWNENASSAGRDMLHYSDDFLCGNGGVPTDCK 113
QY 112 NGFYQYMDPSTGAPAYNDGPTGLQGLDYAIAASAAHGLRVIVLTDNKNKEFGMDQYD- 170
DB 114 -----PGAVVTAFDH--KSLNGCAYSIVT-----LQMGAYVSRD--KNGPVDSE 155
QY 171 ----KWGLPYHDN--FYTDPRQ--QAYKN--WVNHLLNRVNSITGVTKNDPT-----IFA 218
DB 156 APSPRDKVEFAKNAFSLQPLNDQGVYMDDEVNVLNR-----YGNASTSGIRA 207
QY 219 WELANPCRVGSGTLPT--SGTCTOATIVNWDQSAVYKSIDPNHVV-----SVGDGFFYI 273
DB 208 YSLDNEP-ALWSETHPRIHPEQLQALAEVAKSIDLSKAVKNVDPHAEIFCPALYGFAYL 266
QY 274 GSTGSGMP-----YNDPSDGV---DNALLRVKNIDFGTYHLYPNYWG---- 314
DB 267 SLQDAPGWPSLOGNSWFDYDYLQMKNAHTONGKRLDLVDV-----HWYPEAQGGOR 321
QY 315 -----ONADWGTOWIKDHIANAAGIKPTILEEFGWOT 347
DB 322 IVFEGAGNIDTKARVQAPRSLWDPAQEDSWIGTWFSYL-----PLI----- 365
QY 348 PDROSDVYOTW---TQTVRTNGAGW-----NFWMLAGNVN 379
FT SIGNAL 1 31 POTENTIAL.

366 PKLQSSITQYYPGKTLAITESSYCGDNHISGIIATADALGIFGKYGVYAANYMOTEDNTI 425
380 -----COPYPNYDGFNVYPS-----STATVLAS-----EALAISTGTSPP 415
426 YTSAAAYKLYRNYDGNKSGFGSIKVDAATSDTENSSTVYASVYTDENSELHILVLNKNFDDP 485
416 PPSPPSSPS-----SSP 451
486 INATFQLSGDKTYTSGRWGFDOTGSDITQQAATINNNQFTYTLPLSAYHIIVLNKADS 545
452 SSPVSGGVKVOYKND-SAPGDNIQKPLQLVNTGSSVDLSVTVRYWFTTRDGSSTLV 510
546 TEPVNSDLVVOYKGDGRNATDNIQKPHFNINQKGTSPVDLSLTLRYFTFKD-SAAAMN 604
511 YNDWAAMCCNIRASFGSVNPTPATDYLOLSFT--GGTLAAGSGTGELQNRVKNKSDW 568
605 GWTDWAKLGSNTQISFNGINGA--DSDTYAELGFSFGAGSIAEGGSGEIQLRMSKADW 662
569 SNEDETNDYSY-GTNTAFQDQWTKVTVVNGRLVMGTEP 605
663 SNEANEANDYSFDCAKTAYIDWDRVTLQDGLVWGIEP 700

RESULT 13
GUNV_ERWCA STANDARD; PRT; 505 AA.
ID GUNV_ERWCA STANDARD; PRT; 505 AA.
AC Q47056;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
GN CELV.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI193;
RX MEDLINE=94067016; PubMed=8246888;
RA Cooper V.J.C., Salmund G.P.C.;
RT "Molecular analysis of the major cellulase (CelV) of Erwinia
carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
domains".
RL Mol. Gen. Genet. 241:341-350(1993).
CC -!- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH
OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGRFES
CELSIUS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76000; CAAS3592.1; -
DR HSSP; O85465; IAH3.
DR InterPro; IPR001956; CBD_3.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; P50059; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.

	Query Match	9.6%	Score 394:	DB 1:	Length 145:
	Best Local Similarity	50.7%:	Pred. No. 6.9e-13:		
	Matches	75;	Conservative 31:	Mismatches 36:	Indels 6:
				Gaps 4:	
Qy	461	VOYKNNSAFCDNQIKPGLLOLVNTGSSVDLSITVTVYVWTFTRDGGCSSTI.VYNCDDWAAMCIC	520		
Dd	1	LOYRAADTAADNOIQKPFNIKNGTSAVDLSLTIRYFTFKDGSAAVNCM-IDMAQLGG	54		
Qy	521	GNIRASFGSNVPAPTADTVLQISFT--GGTLAAGSGTGFEIQNRVKNSDKWSNFDNDYS	578		
Dd	60	SNIQISFG--NHCTGTSNTDTVVELSFSFSEACGSIAAGCGSETQLRMSKTDSNFNENDYS	117		
Qy	579	Y-GRTNAFAQDWTKVTVYVNGRLVNGTEP	605		
Dd	118	FDGTTKTAFAQMDWRVVLYONGIVNGTAP	145		

RESULT 15	
GUNA_MICBI	
ID	GUNA_MICBI
AC	STANDARD;
DT	PRT;
DT	456 AA.
DT	P26414;
DT	01-AUG-1992 (Rel. 23, Created)
DT	01-AUG-1992 (Rel. 23, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	Endoglucanase A Precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE	(Cellulase).
GN	CELA.
OS	Microbisporea bisporea.
OC	Bacteria; Actinobacteria; Actinobacteriia (Class); Actinobacteridae;
OC	Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
OC	Thermobispora.
OX	NCBI_TaxID=2006;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Yablonsky M.D., Elliston K.O., Eveleigh D.E.;
RL	(In) Coughlan M.P. (eds.);
RL	Production, characterization and application of cellulose,
RL	hemicellulose and lignin enzyme degrading systems, pp.77-83, Elsevier,
RL	London (1989).
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC	linkages in cellulose.
CC	-1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC	(CBD).
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC	HYDROLASES).
CC	HSPSP; P26222; 1TML.
DR	InterPro: IPR001919; Bac_celose-bind.
DR	InterPro: IPR001524; GH_6.
DR	Pfam: PF00553; CBM_2; 1.
DR	Pfam: PF01341; Glyco_hydro_6; 1.
DR	PRINTS; PR00733; GLHYDRLASE6.
DR	PRODOM; PD003733; GH_6; 1.
DR	PROSITE; PS00561; CBD_BACTERIAL; 1.
DR	PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR	PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT	SIGNAL
FT	CHAIN
FT	31 456
FT	ENDOGJUCANASE A.

FT	DOMAIN	31	322	CATALYTIC.
FT	DOMAIN	323	355	LINKER ("HINGE") (PRO-SER BOX).
FT	DOMAIN	356	456	CELLULOSE-BINDING.
FT	ACT_SITE	113	113	BY SIMILARITY.
FT	ACT_SITE	151	151	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	300	300	NUCLEOPHILE (BY SIMILARITY).
FT	DISULFID	114	159	BY SIMILARITY.
FT	DISULFID	267	302	BY SIMILARITY.
FT	DISULFID	360	453	BY SIMILARITY.
SQ	SEQUENCE	456 AA;	47011 MW;	B06D8595E322848F CMC64;
Query Match				
Best Local Similarity 9.28; Score 380.5; DB 1; Length 456;				
Matches 144; Conservative 63; Mismatches 206; Indels 113; Gaps 24;				
Qy	255	VKSIDPNHWSVGDGCFYIG-STQSGWPNYPNDPSQVONNALLRVKNIDFGTYVHLYPNYW	313	
Db	26	IASAGPAIAY---DSPFYVDPOSNAAKWVAANPND--PRTPVIRDIRIAAVPTGRWFANY-	79	
Qy	314	QONADWGTMWKDHTANAAAGKPTILEEFQWQTPDRDSVYQVTWTQTVRTNGEAGWNFMW	373	
Db	80	--NPSTVRAEVDAYVGAAGAAACKIPIMVYA--MPNRD-----	113	
Qy	374	LAGNYNGQYPNYDGFNYYYFSSTATVLASEALAISTGTPPPSPSSSPSSPSPSPSPS	433	
Db	114	-CGGPSAGGAPNITAYRAWIDEIAAGLRNRPVAIL-----EPDALPIMTNCMSPSEQ	165	
Qy	434	ASPSASPSASSP	489	
Db	166	AEVOASAVGACKKPKAASSQ-----AKVYFDAGHDWVPADEMASRLRGADIANS-ADGI	219	
Qy	490	DLSTVTVRYWFTRDGGSSTLYV-NCDMAAMCGNTRASFQ-SVNPATPTADTYLQLSETG	547	
Db	220	ALNVSNYRY-----TSGLISYAKSVLSAIGASHLRRAVIDTSRNGNGPLGSEWCD-----	268	
Qy	548	CTLAAGGSTGEIQNRVKNKSDKSNDETDNDYSYGTNTAFQDWTKVT-----VYVN	596	
Db	269	---PPCRATG-----TWSTTD-TGDPAL---DAFL-WIKPPGEADGCIATPGVFVP	311	
Qy	597	GRVWCTEPTSP	656	
Db	312	DR---AYELANNAAPPTYSP	355	
Qy	657	SVSSGVCGRATYVVVNSDWGSGFTATVTVNTGSRATSGWTVAMSGGNQTVNTYNTAL	716	
Db	356	----AGRAACEATYALVNWMPGCGFOAEVTVKNTGSSPINGWTVQWTLPSGQSITQLWNGDL	411	
Qy	717	TQSGASVTATNLSYNNVLOPGOSTTFGFGNSYSGTNTAPTLLTCTAS	762	
Db	412	STSGSNVTVRNWSNGNVNPGAGSTSGFLGSGTG-QLSSSITCSAS	456	

Search completed: November 13, 2002, 11:55:00
Job time : 22.5172 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:50:41 ; Search time 70.6164 Seconds
(without alignments)
2223.392 Million cell updates/sec

Title: US-09-917-378-1

Perfect score: 4122
Sequence: 1 MGLVRRPARAFVATAACTAV.....GFNGSYSGTNTAPTTLCTAS 762

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 205047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_protist.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	841.5	20.4	666	2 Q9RIK7	Q9rik7 thermotoga
2	813	19.7	680	2 Q9RIK9	Q9rik9 thermotoga
3	811	19.7	669	16 Q9X0V4	Q9x0v4 thermotoga
4	790	19.2	694	2 Q9Z187	Q9z187 bacillus st
5	740	18.0	921	2 Q9L8L8	Q9l8l8 caldicellul
6	723.5	17.6	439	3 Q9Z401	Q9z401 agericus bi
7	719.5	17.5	439	3 Q9P893	Q9p893 agericus bi
8	684.5	16.6	930	2 Q9RFX5	Q9rfx5 caldicellul
9	664.5	16.1	437	3 Q99036	Q99036 trichoderma
10	661.5	16.0	1711	2 P96311	P96311 anaerocellu
11	644	15.6	377	3 Q90012	Q90012 aspergillus
12	624	15.1	1751	3 Q9A0G4	Q9a0g4 caldicellul
13	619	15.0	578	3 Q96V96	Q96v96 orpinomyces
14	609.5	14.8	1000	2 Q24820	Q24820 thermophili
15	609	14.8	1779	2 Q52374	Q52374 caldicellul
16	602.5	14.6	1770	2 Q9X3P5	Q9x3p5 caldicellul

17	589.5	14.3	1426	2	Q9X3P6	Q9x3p6 caldicellul
18	552	13.4	431	10	Q9FJ23	Q9fj23 arabidopsis
19	544.5	13.2	416	10	Q9FT03	Q9ft03 coffea arab
20	541	13.1	408	10	Q9SG95	Q9sg95 arabidopsis
21	532	12.9	395	10	Q8RVL3	Q8rvl3 lycopersico
22	530	12.9	399	10	Q93WT4	Q93wt4 lycopersico
23	526.5	12.8	411	10	Q9FZ29	Q9fz29 arabidopsis
24	519	12.6	414	10	Q9SG94	Q9sg94 arabidopsis
25	498.5	12.1	996	2	Q9AQI0	Q9aqi0 caldicellul
26	498	12.1	448	10	Q9LZV3	Q9lzv3 arabidopsis
27	498	12.1	468	10	Q8SAY1	Q8say1 oryza sativ
28	495	12.0	431	10	Q9M0H6	Q9m0h6 arabidopsis
29	494	12.0	427	10	Q9FT02	Q9ft02 coffea arab
30	491.5	11.9	439	10	Q8SAY2	Q8say2 oryza sativ
31	491.5	11.9	445	10	Q94J47	Q94j47 oryza sativ
32	489.5	11.9	170	2	Q9RFX6	Q9rfx6 caldicellul
33	466.5	11.3	415	10	Q9FZ03	Q9fz03 lycopersico
34	464	11.3	401	10	Q9FU06	Q9fu06 lycopersico
35	463	11.2	499	2	Q93TJ6	Q93tj6 bacillus su
36	458	11.1	508	2	Q93LD0	Q93ld0 bacillus su
37	451.5	11.0	499	2	Q45532	Q45532 bacillus su
38	450	10.9	499	2	Q52731	Q52731 bacillus sp
39	433	10.5	1915	2	Q9KPL0	Q9kpl0 acetivibrio
40	430.5	10.4	501	2	Q83012	Q83012 bacillus sp
41	427.5	10.4	997	2	Q9Z411	Q9z411 bacillus sp
42	426.5	10.3	369	10	Q48540	Q48540 lycopersico
43	423.5	10.3	442	16	Q9AA10	Q9aa10 caulobacter
44	415.5	10.1	403	10	Q9SKU9	Q9sku9 arabidopsis
45	407	9.9	311	10	Q9SLW4	Q9slw4 arabidopsis

ALIGNMENTS

RESULT 1

Q9RIK7 ID Q9RIK7 PRELIMINARY; PRT: 666 AA.
AC Q9RIK7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Beta-mannosidase (Beta-mannanase).
GN MANB OR MAN5.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2337;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5068;
RA Parker K.N., Lam D., Duffaud G., Snead M.A., Mathur E.J., Kelly R.M.;
RT "Amino Acid Sequence of beta-mannosidase genes from the
RT hyperthermophilic bacteria Thermotoga maritima and Thermotoga
RT neapolitana";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 5068;
RA PARKER K.N., CHAHBRA S.R., LAM D., CALLEN W., DUFFAUD G.D.,
RA SNEAD M.A., SHORT J.M., MATHUR E.J., KELLY R.M.;
RT "Galactomannanases Man2 and Man5 from Thermotoga species: growth
RT physiology on galactomannans, gene sequence analysis, and biochemical
RT properties of recombinant enzymes";
RL Biotechnol. Bioeng. 75:322-333(2001).
DR EMBL; Y1981; CAB56856.1; -;
DR EMBL; AY033477; AAK53459.1; -;
DR InterPro; IPR001547; GH_5.
DR Pfam; PF00150; cellulase_1.
SQ SEQUENCE 666 AA: 76695 MW: 7EF986115E4C58CD CRC64;

Query Match 20.4%; Score 841.5; DB 2; Length 666;
Best Local Similarity 43.1%; Pred. No. 6.6e-35;
Matches 175; Conservative 65; Mismatches 123; Indels 43; Gaps 13;

```

:
:
QY 29 ITMPSATAAPAGFVTASGGQFVLNGLPYRYGCTNNYYLSYQSHADVDVYLAKAQAMNLSV 88
Db 11 VTLPLVLFSNDSFVKVNGCRFLJNGEEFRFCVGSNNYMHYKSNRMIDSVLESAGKMGVKV 70
QY 89 IRTWCFIDIGSLDGSVPITDGNKNGFYQYWDPSGTA---PAYNDGPTGLOGLDYATASA 145
Db 71 LRWNGF-----LDGESYCRDKN-----TYMHPEPGVFCGLPECTNAQDGFERLDYTVAKA 119
QY 146 AAHLGRVIVLTLNDWKKEFGMDQYDKWYGLPYHDNFYTDPTQOAYKNNVHLLNRVNSI 205
Db 120 KELGKLIIVLNNWDDFCGMNQYVRWEGGIHDDFYRNEKIKEEYKVKYVSLINRVNTY 179
QY 206 TCVTYKNDPTIFAWELANPRCVGSCTLPTSGCTCTOATIVNVWDQMSAYVKSIDPNHMS 265
Db 180 TCVPTREPTIFAWELANPRC-----ETDKSGN-----TLVEWVEEMSAVIKSLDPNHLVA 231
QY 266 VGDEGF---YIGSTQSG---WPYNDPSDGVNALLRVKNIDFGTYIILPYNWQGN--- 316
Db 232 VGDEGFNNYEGFRPYGGEAEWAYNGWS-GVDMKKLLEIETVDFCTFHLPSHWGVS PEN 290
QY 317 -ADWCTOWIKDHIAAAGKPTILEEFG--WOTP-DRDSVYQVTVQVTVRTNGEAGWNFW 372
Db 291 YAQWAGAKWIEDHIIKAKEIGKPVVLEEYGIKPSAPVNRVAIYKLNWDLVYNLGGNCAMFW 350
QY 373 MLAGNVNGOP-----YPNYDGFNVVYPSSTATVLASE-ALAIISG 411
Db 351 MLAGICEGDRDEKGYYPDYDGFRIVNDSESAKLIJREYAKLFFSTG 396

RESULT 2
QYRIK9 PRELIMINARY: PRT: 680 AA.
ID QYRIK9
AC QYRIK9:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Beta-mannosidase.
GN MANB.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSR8;
RA Parker K.N., Lam D., Duffaud G., Snead M.A., Mathur E.J., Kelly R.M.;
RT "Amino Acid sequence of beta-mannosidase genes from the
RT hyperthermophilic bacteria Thermotoga maritima and Thermotoga
RT neopolitana.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y17980; CAB56854.1; -.
DR InterPro: IPR004200; Bgal_small_C.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF02930; Bgal_small_C; 1.
DR Pfam: PF00150; cellulase; 1.
SQ SEQUENCE 680 AA: 77687 MW: 58A3F6982194D1AA CRC64:

Query Match 19.7%; Score 813; DB 2; Length 680;
Best Local Similarity 41.5%; Pred. No. 1.8e-33;
Matches 173; Conservative 65; Mismatches 135; Indels 44; Gaps 13;

QY 19 AVAAATLGSITMPSATAAPAGFVTASGGQFVLNGLPYRYGCTNNYYLSYQSHADVDVYL 78
Db 12 SVSABFLLLIVELSFVLFASDFVKVNGKFKALNGKEFRFGICSNYYMHYKSNMGIDSVL 71
QY 79 AKAAAMNLSVIRTWGTFIDIGSLDGSVPTIDGNKNGFYQYWDPSGTAAYNDG----PTG 134
Db 72 ESARDMGIKVLWNGF-----LDGESYCRDKN-----TYMHPEPGVFCVPEGISNAOSG 120
QY 135 LOGLDYATASAAHGLRVIVLTLNDWKKEFGMDQYDKWYGLPYHDNFYTDPTQOAYKNNW 194
Db 121 FERLDYTVAKAKELGIKLVILVNNWDDFGGMNQYVRWEGGTHDDFYRDEKIKEEYKY 180
```

```

QY 195 VNHLLNRVNSITGVYKNDPTIFAWELANPRCVSGTLPTSGCTCTOATIVNVWDQMSAY 254
Db 181 VSFLVNHVNTYGVYRREPTIFAWELANPRC-----ETDKSGN-----TLVEWVKEMSSY 232
QY 255 VKSIDPNHMSVGDEGCF---YIGSTQSG---WPYNDPSDGVNALLRVKNIDFGTYIHL 308
Db 233 IKSLDPNHLVAVGDEGFFSYEGFKPYGGEAEWAYNGWS-GVDMKKLLSIETVDFCTFHL 291
QY 309 YPNYWQGN-----ADWCTOWIKDHIAAAGKPTILEEFG--WOTP-DRDSVYQVTVQV 361
Db 292 YPSHWGVSPEYTAQWAGAKWIEDHIIKAKEIGKPVVLEEYGIKPSAPVNRVAIYKLNWDLV 351
QY 362 RTNGEAGNFWMLAGNVNGOP-----YPNYDGFNVVYPSSTATVLASE-ALAIISG 411
Db 352 YDLGGDGMFWLAGICEGDRDEKGYYPDYDGFRIVNDSDPEALIRYAKLFFNTG 408

RESULT 3
QYXOV4 PRELIMINARY: PRT: 669 AA.
ID QYXOV4
AC QYXOV4:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Endo-1,4-beta-mannosidase.
GN TM1227.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSR8 / DSM 3109;
RX MEDLINE=99287316; Pubmed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL: AE001779; AAD36302.1; -.
DR TIGR: TM1227; -.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
KW Complete proteome.
SQ SEQUENCE 669 AA: 76931 MW: 5659B727305688F5 CRC64:

Query Match 19.7%; Score 811; DB 16; Length 669;
Best Local Similarity 42.8%; Pred. No. 2.3e-33;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

QY 41 FVTASGGQFVLNGLPYRYGCTNNYYLSYQSHADVDVYLAKAQAMNLSVIRTWGTFIDIGSL 100
Db 23 FVKVNGKFKALNGKEFRFGICSNYYMHYKSNRMIDSVLESARDMGIKVLRIWNGF-----L 77
QY 101 DCSVPTIDGNKNGFYQYWDPSGTAAYNDG----PTGLOGLDYATASAAHGLRVIVL 156
Db 78 DECSYCRDKN-----TYMHPEPGVFCVPEGISNAOSGFERLDYTVAKAKELGIKLVIL 131
QY 157 TNDWKEFGMDQYDKWYGLPYHDNFYTDPTQOAYKNNVHLLNRVNSITGVYKNDPTI 216
Db 132 VNVWDDFGGMNQYVRWFGTGHDDFYRDEKIKEEYKVKYVSLVNHVNTYTGVPYRREPTI 191
QY 217 FAWELANPRCVGSCTLPTSGCTCTOATIVNVWDQMSAYVKSIDPNHMSVGDEGCF---YI 273
Db 192 MAWELANPRC-----ETDKSGN-----TLVEWVKEMSSYIKSLDPNHLVAVGDEGFSNYE 243
QY 274 GSTQSG---WPYNDPSDGVNALLRVKNIDFGTYIHLVYPNWQGN-----ADWCTOWIKD 326
Db 244 GPKPYGGEAEWAYNGWS-GVDMKKLLSIETVDFGTFHLPSHWGVSPEYTAQWAGAKWIED 302
```


Db 127 SDIANAGGTTVTRWGFNEVTS-----PNCNYYQSW---SGARPTINTGASGLI.NF 173

Qy 139 DYAIASAAAGLRVIVLVLTNDWKEFGGMDQYDKWY---GLPYHDNFYTDPRTOQAYKNWV 195

Db 174 DNVIAAKANGIRLJIALVTNNWADYCGMDVYVNMWCGQP-HDLFYNTAIAKDAPKSYG 232

Qy 196 NHLLNRVNSITGVVTKNDPTIFANELANERPCVSGSGLTPTSGCTGCTQATIVNWWDNSAYV 255

Db 233 RAFVSR-----YANEPTVMAMELANERPRCKGS-TCITTSCTCTTTTNTVMAKENSAFI 283

Qy 256 KSIDPNHMVSGDECFYIGSTOGSCWPNNDPSDCVDNNALLRVKNIDFGYTHLYPNYMQQ 315

Db 284 KTIDSNHLVALIGDEGFY-NOPGAPTYDQG-SEGVDFEANLAISSVDFAFHISTPEPMGQ 341

Qy 316 NAD---WCTOWIKOIHANAAAIKRPITILEEFGWQTPORDSVYQYOTWTQVTFNTEAGNFW 372

Db 342 CADAKANGTQWITDHAASMKRVNKPVILEEFGVTTNQPDY-VAEWFNIESSGLTGDLIW 400

Qy 373 MLCNVNGQPPNVYDGFNVYPSSTATVLAASEALA1 408

Db 401 QAGSHLSTGDTPN-DGYAVYDGPVYPLVKSHASAM 435

RESULT 7

Q9P893 PRELIMINARY; PRT; 439 AA.

ID Q9P893

AC Q9P893;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE CEL4a mannannase precursor (EC 3.2.1.78).

GN CEL4.

OS Agaricus bisporus (Common mushroom).

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

OC Agaricales; Agaricaceae; Agaricus.

OX NCBI_Taxid=5341;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C54-CARB8; TISSUE=MYCELIUM;

RE MEDLINE=21218546; PubMed=11319115;

RA Tang C.M., Waterman L.D., Smith M.H., Thurston C.F.:

RT "The cel4 gene of Agaricus bisporus encodes a beta-mannanase."

RL Appl. Environ. Microbiol. 67:2298-2303(2001).

DR EMBL; AJ271862; CAB76904.1; -

DR HSSP; P00725; IAZ6.

DR InterPro; IPR001589; Actbind_actnin.

DR InterPro; IPR000254; CBD_fungal.

DR InterPro; IPR001547; GH_5.

DR Pfam; PF00734; CBM_1; 1.

DR Pfam; PF00150; cellulase; 1.

DR ProDom; PD001821; CBD_fungal; 1.

DR SMART; SM00236; fcbd; 1.

DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.

DR PROSITE; PS00562; CBD_FUNGAL; 1.

KW Glycosidase; Hydrolase; Signal.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 439 CEL4A MANNANASE.

FT SEQUENCE 439 AA; 47042 MW; F11885A4BD9591CC CRC64;

Query Match 17.5%; Score 719.5; DB 3; Length 439;

Best Local Similarity 40.6%; Pred. No. 5.8e-29;

Matches 160; Conservative 57; Mismatches 140; Indels 37; Gaps

Qy 25 TLGSIWPSATAAP--AGFVTASGQFVLNGLPYRGCTNNY--LSYOSHADVDDVIAK 80

Db 69 TTSOTTAPPTTSHPVTFGVKASGTRFLTNQKQYTVVGGNSYVWGLTSLSAMNOAFSD 128

Qy 81 AQAMNLNVIRTWGFDIGSLDGSVPTIDGNKNKGFYFOYWDPSGTGAPAYNDGPTGLOGLDY 140

Db 129 IANAGGTTVTRWGFNEVTS-----PNCNYYQSW---SGARPTINTGASGLI.NF 173

Qy 141 AIASAAAGLRVIVLVLTNDWKEFGGMDQYDKWY---GLPYHDNFYTDPRTOQAYKNWNI 197

```
Db 176 VIAAKANGIRLIVALTNNWADYGGMDVYVYVQWVGNQOP-HDLFYNTAATKDAFKSVYRT 234
QY 198 LLNRVANSITCVTYKNDPTIFAMELANEPRCVGSGTLPTSGCTQATITVNVMDVAVVKS 257
Db 235 FVSR-----YANEPTVAMELANEPRCKGS-TGTTCTCTTIVTNWAKEMSAFIKT 285
QY 258 IDPNHVSVDGEGFYLGSGGWPYNDSDGVNDNALLRVKNDIDFGTYHLYPNYWGQNA 317
Db 286 IDSNNHLVAIGDEGFI-NQPCAPTYQGG-SEGVDFAEALAISSVDFAFHSYPEPQGA 343
QY 318 D---MCTQWIKDHIAAAAIKGPRTILEEFGWQTPDRDSVYQVWTQTVRTNGEAGWFML 374
Db 344 DAKANGTQWITDHAASKRVNKPVILEEFGVTTNQPDY-VAEFNVESSGLTGDLIMQA 402
QY 375 AGNVGOPYNVDCGFVNYVYSSSTATVLASEALAI 408
Db 403 GSHLSTGDTN-DCAVAYPDGPVYPLMKSHASAM 435

RESULT 8
Q9RFX5
ID Q9RFX5 PRELIMINARY: PRT: 930 AA.
AC Q9RFX5;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN MANA.
OS Caldicellulosus cellulosovorans.
OC Bacteria; Firmicutes; Bacillales; Clostridium group; Bacillales;
OC Alicyclobacillaceae; Caldicellulosus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosus cellulosovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AACF22274.1; .
DR HSP; O06831; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR004302; Chitin_binding_3.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00942; CBM_3; 2.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF03067; Chitin_bind_3; 1.
DR PRINTS; PRO1217; PRICHEXTENS.
DR ProDom; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 33 POTENTIAL.
FT CHAIN 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
SQ SEQUENCE 930 AA: 101576 MW: 0086638D54D1A2CC CRC64;

Query Match 16.6%; Score 684.5; DB 2: Length 930;
Best Local Similarity 30.9%; Pred. No. 7.7e-27;
Matches 168; Conservative 96; Mismatches 149; Indels 131; Gaps 18;

QY 304 GYHLYPNYWGQADNGTQWIKDHIAAAAIKGPRTILE-----EFGWQ-----PDRDS 352
Db 143 GTWLYVTVDGDPDTPQLKWSOLEPTFPFQVNTNPPINSSGPDCAEYSWQVQLPNKQGRHI 202
QY 353 VYQVTTQTVRTNGEAGWFWMLAGNVNGOPYPNYDGFVNY-----SDSPEAFYNCSDVYFGSGPIAYEFGDPREGGTMI 401
Db 203 IYMIQR-----SDSPEAFYNCSDVYFGSGPIAYEFGDPREGGTMI 243
QY 402 ASEALAIATGTPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 461
Db 244 TPP-----PSGTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 492
```

```
QY 462 QYKNDSAPGDNOIKPOLQLVNTGSSSVLDLSTVTYRYWFTRDGSGSTLVYNCDMAAGCG 521
Db 293 EYRVGDTSATDNQMKPOLRVNTGSOAVPLTELKRVYWTKN-STOAEQYFCDMAOIGCS 351
QY 522 NIRASGCVNPAITADTYLQLSFTGCTLAAGSGTGIEQNRVKNKSDNSNDETDNDYSY-C 580
Db 352 NIRAQFVLSQVSGADSYIELSFTGSI PAGGTGIEQNRHIIHFTNMNMYNETDDSYNG 411
QY 581 TMTAFODWTKVTYVNCRLVMGTPEPSTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 640
Db 412 AQTWGPSTRTILYRNCVLVMGTPECGSGSPPTPTPTPT-----PTPTPTPTPTPTPTPT 465
QY 641 SPSPSTSPSPSPSPSP-----SVSSSGV-----GCRATY 669
Db 466 TSTPTSGGNLNVNTQGLVGINHPAWRDLRLSSSLQIRSMCANAVRIVLSNGCRWTK 525
QY 670 VVNSDMGSGFTATVTVTNTGSR-----TSWTVAVSFGGNOTVTNYW----- 712
Db 526 IPASEVADIISQARTL---GYRAVLEVHDTTGYGEDAAACSMTTAVNYVIELKNVLQ 582
QY 713 -NTALTSQSGAS-----VTAT-----NLSYNNVIOPCOSTTFCFNGSYSCNT 753
Db 583 ENFVIVNIGNEPYGNVYQNVWDTDRNAVALRAGINNTIMV-DAPNMGDWSFTMKDN 641
QY 754 APTL 757
Db 642 APTI 645

RESULT 9
Q99036
ID Q99036 PRELIMINARY: PRT: 437 AA.
AC Q99036;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Beta-mannanase precursor.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RUT30;
RA Stahlbrand H., Saloheimo A., Velmaannpera J., Penttila M.;
RT "CDNA encoding Trichoderma reesei beta-mannanase.";
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L25310; AAA34208.1; .
DR HSSP; P00725; IAZ6.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00734; CBM_1; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 19 POTENTIAL.
FT CHAIN 28 437 BETA-MANNANASE.
SQ SEQUENCE 437 AA: 47053 MW: 17513DADEI2654A7 CRC64;

Query Match 16.1%; Score 664.5; DB 3: Length 437;
Best Local Similarity 35.3%; Pred. No. 3.4e-26;
Matches 157; Conservative 73; Mismatches 162; Indels 53; Gaps 16;

QY 18 TAVAAAATLGSITMPSATAAPAGFTVATAGGQFVNLGFLPYRGCTNNYYSY-OSHADVDD 76
Db 10 SAATAAALAAQLQVPRA--SSFTISGTQFNIDKVGVFAGTNCYWCSTLNHADVDS 67
QY 77 VLAKAAMNLVIRTWGFIIDIGSLDGSVPTIDGNKNGFYQYWDPSGAPAYNDGPTGLQ 136
Db 68 TFSHTISSGCLKVVVRWGFNDVNT-----QPSQGIWFOKLS-ATGS-TINTGADGLQ 117
QY 137 GLDYAIAASAAHGLRIVVLTNDWKEFGMDQDKWYGLPYHDNFYTDPTQQAAYKNVNV 196
```


Db	118	TLDYVQSSAEQINLKLIIPFVNNWSDYGGINAYNAFG-GNATTWYNTAAQTQYRYVQ	176			
Qy	197	HLNLRVNSITGYTKNDPTIFAWELANPRCVGSGTLLPTSGCTCTQATIVNVNVDMSAYVK	256			
Db	177	AVYSR-----YANSTAFIAWELGNPCNG-----CSTDVIVQWATSVSOYVK	219			
Qy	257	SIDPNHMYSGDEGYIGSTGSG-WPYNDRSDGDVNNALLRVKNIDFGYHLYPNYWGQ	315			
Db	220	SLDSNHLVTLGDC--LGLSTGDGAYPT-YCECTDFAKNVQIKSLDFGTFHLYPDSMGT	276			
Qy	316	NADWCTQWIKDIIANAAAIKRPITILEEFG-WQTP-DRDSVYQTWQTQVTRNGEAGNFWM	373			
Db	277	NYTWGNGLIQTAAACLAAGKPCVEEYGAQNPCTNEAPWQTISLT--TRMGCDMFWQ	334			
Qy	374	LACNVGQPYPNYDENVYSPSTATVLAESAL-AISTGTSPSPSPSSSPSSPSPSPSP	432			
Db	335	WGDTFANGAQSNDPYTWYNSNMOCVLKHNVAINGITTPPPVSTTTTS-----	387			
Qy	433	SASPSASPSASPSPPSPSPSPSG 457				
Db	388	-----SRTSTPPPPGSCSPLYG 406				
RESULT 10						
P96311						
ID	P96311	PRELIMINARY; PRT; 1711 AA.				
OC	P96311					
DT	01-MAY-1997	(TReMBLrel. 03, Created)				
DT	01-MAY-1997	(TReMBLrel. 03, Last sequence update)				
DT	01-JUN-2002	(TReMBLrel. 21, Last annotation update)				
DE	Endoglucanase A (EC 3.2.1.4) (Endo-1,4-beta-glucanase A) (Cellulase A) (Fragment).					
DE	CELA.					
GN	Anaerocellum thermophilum.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;					
OC	Clostridiales; Anaerocellum group; Anaerocellum.					
OX	NCBI_Taxid=31899;					
RP	[1]					
RN	SEQUENCE FROM N. A.					
RC	STRAIN=Z-1320;					
RA	Zverlov V.;					
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.					
CC	-!- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN					
CC	ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCULOSE. THE C-TERMINAL					
CC	DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE					
CC	CELLULOSE (BY SIMILARITY).					
CC	-!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC					
CC	LINKAGES IN CELLULOSE.					
CC	-!- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULASE FAMILY E					
CC	-!- (FAMILY 9 OF GLYCOSYL HYDROLASES).					
CC	-!- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULASE FAMILY L					
CC	(FAMILY 48 OF GLYCOSYL HYDROLASES).					
EMBL	EMBL; Z86105; CAB06786.1; -.					
HSSP	HSSP; P26221; ITP4.					
DR	InterPro; IPR001956; CBD_3.					
DR	InterPro; IPR001701; GH_9.					
DR	InterPro; IPR000556; Glyco_hydro_48.					
DR	Pfam; PF00942; CBM_3; 3.					
DR	Pfam; PF02011; Glyco_hydro_48; 1.					
DR	Pfam; PF00759; Glyco_hydro_9; 1.					
DR	TrnTMS; PR00844; GLHYDRASE48.					
DR	ProDom; PD001947; CBD_3; 2.					
DR	ProDom; PD011903; Glyco_hydro_48; 1.					
DR	KOSI7B; PS00592; GLYCOSYL_HYDROL_F9_1; 1.					
DR	PROSITE; PS00698; GLYCOSYL_HYDROL_F2; 1.					
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.					
DR	NON_TER 1					
DR	SEQUENCE 1711 AA; 189979 MW; E3E987CEB9C00C21 CRC64;					
Query Match 16.08; Score 661.5; DB 2; Length 1711;						
Best Local Similarity 25.18; Pred. No. 2.2e-25;						
Matches 222; Conservative 121; Mismatches 245; Indels 295; Gaps						

[illegible]

Query Match	16.0%	Score 661.5;	DB 2;	Length 1711;	DE Mannanase precursor.
Best Local Similarity	25.1%	Pred. No. 2.2e-25;			GN MAN1.
Matches 222;	Conservative 121;	Mismatches 245;	Indels 295;	Gaps 36;	OS Aspergillus aculeatus.
					OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX	NCBI_TaxID=5053;
RN	[1]
RC	SEQUENCE FROM N.A.
RP	STRAIN=KSM 510;
RX	MEDLINE=95078777; PubMed=7987261;
RA	Christgau S., Kauppinen S., Vind J., Kofod L.V., Dalboge H.;
RT	"Expression cloning, purification and characterization of a beta-1,4-
RL	mannanase from Aspergillus aculeatus.";
RR	Biochem. Mol. Biol. Int. 33:917-925(1994).
DR	EMBL: L35487; AAA67426.1. -
DR	InterPro: IPR001547; GH5.
DR	Pfam: PF00150; cellulase; 1.
KW	Signal.
FT	SIGNAL 1 18 POTENTIAL.
FT	CHAIN 19 377 MANNANASE.
FT	SEQUENCE 377 AA; 41082 MW; 10E6477553E3CA2 CRC64;
QY	Query Match 15.6%; Score 644; DB 3; Length 377;
QY	Host Local Similarity 35.2%; Pred. No. 3.1e-25;
QY	Matches 138; Conservative 67; Mismatches 147; Indels 40; Gaps
DB	24 ATGCSIT----MPSATAAPACGFVTASGGQVLNGLPYRYGGTNNYLSY-QSHADYDDVL 78
DB	11 ASLCVATALPRTPHNNAATTAFTSTSLCHFTIDGKTCYFAGTNSYWGELTNDDVDLVM 70
QY	79 AKQAAMNLSVIRTWGFTDIGS--LDGSVPTIDGNKNGFFQYWDPSGTGAPAYNDGPTGLQ 136
DB	71 SOLAASDLKTLRWYGVNDVNTKPTDGT-----WYQLHANGTSTINTGADGLQ 118
QY	137 GLDYATASAAHGLRVIVLTNDWKEFGGMDQYDKWGLPYHDNFYTDPRTOOAYKNWYN 196
DB	119 RLDYVVTSAEKYGVKLIINFVNEWTYDGMQAVTAYGAAATDFVTNTAIOAYKNYIK 178
QY	197 HLLNRVNSITGVYKNDPTIFANELANEPKCVSGTLPTSGTCTQATIVNWVDOMSAYVK 256
DB	179 AVVSR-----YSSAAIFANELANEPKCG-----CDTSVLNWNISDTSKYIK 221
QY	257 SIDPNHNVSYDGEFGYIGSTQSGWPNPDSGVDNALLRVKNIDFGTYHLYPNYWGON 316
DB	222 SLQSKHLVTIGDEFGGLDVSDSGSYPT-YGEGLNFTKNLGISTIDFGTLHLYPDSWGTS 280
QY	317 ADNGTOWIKDHIAANAAIGKPTILEEFGMTDPDRSDSVYQTWTQTV-RTNCEAGWNFWMLA 375
DB	281 YDWGNGWITAHAAACKAVGKPCLEEYGV-VTSNHCAVESPWQQTACGNATGISDLYWQYV 339
QY	376 GNVNGQPYNYDGFNVYPSSTATVLAEEALA 407
DB	340 TTTSWGQSPN-DGNTFFYNTSDFTCLVTDHVA 370
RESULT 12	
Q9AQG4	
ID	Q9AQG4 PRELIMINARY; PRT: 1751 AA.
AC	Q9AQG4;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Cele.
OS	Caldicellulosiruptor sp. Tok7B.1.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC	Clostridiales; Syntrophomonadaceae; Caldicellulosiruptor.
OX	NCBI_TaxID=80339;
RN	[1]
RC	SEQUENCE FROM N.A.
RP	STRAIN=Tok7B.1;
RX	MEDLINE=201711169; PubMed=10706665;
RA	Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA	Bergquist P.L.;
RT	"Multidomain and multifunctional glycosyl hydrolases from the extreme
RT	thermophile Caldicellulosiruptor isolate Tok7B.1.";
RR	Curr. Microbiol. 40:333-340(2000).
DR	EMBL: AF078042; AAK06394.1. -

[illegible]

```
QY 694 SGWTVAMS-----FCGNOTVTNYMNTALTOSCATVTATNLSYNNV---IQPGQ 738
DB 936 QSVSVASSINPAYIDVKFKVIGKANAGGADYYVEIGFKSGAGVLAQOSTKEIRLSIQKG- 994
QY 739 STTFGFGNSYSGTN 752
DB 995 -----SGSYNQSN 1002

RESULT 13
Q96V96
ID Q96V96 PRELIMINARY: PRT: 578 AA.
AC Q96V96;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Mannanase Mana.
GN MANA.
OS Orpinomyces sp. PC-2.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimastidaceae; Orpinomyces.
OX NCBI_TaxID=50059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC-2;
RX MEDLINE=21405733; PubMed=11514516;
RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
RA Ljungdahl L.G., Op Den Camp H.J.M.;
RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
RL J. Bacteriol. 183:5325-5333(2001).
DR EMBL: AF177206; AAL01213.1; -.
DR InterPro: IPR002883; CBD_5.
DR Pfam: PF02013; CBM_10; 2.
SQ SEQUENCE 578 AA; 64425 MW; 482A0F36D36F6F66 CRC64;

Query Match 15.0%; Score 619; DB 3; Length 578;
Best Local Similarity 28.6%; Pred. No. 9e-24;
Matches 173; Conservative 76; Mismatches 203; Indels 152; Gaps 21;

QY 29 ITMPSATAAP--AGVVTASGGQVFLNGLPYRYGGTNNYLSYQSHADVDVLAKAQMNL 86
DB 11 LFLTASVACQFREGFVQDTGTFVVDGCKRYFGSNTYYLMVSNHERVDLALETYARHLL 70
QY 87 SVIRTWGFTD-----ICSLDGSVPTIDGNKNGFYFYQWDPSTCAPYNDGPTGLOGLD 139
DB 71 NVVRAWAFCECEDATRLVDFSGPEVLNGE-----NWEKVD 107
QY 140 YAIASAAHGLRVIVLTDNWKKEFGMDQYDKWYGLPYHDNYTDPRTQOAYKNVNHLL 199
DB 108 YYLAAAQNRNIRVLTLNNWTDYGGMDVWVKQFGCKYHDEYTNKDIJKYKQVIKAM1 167
QY 200 NRVNSITGVYKNDPTIFAWELANPRC-VGSGTLPTSGCTCTQATIVNNVDDMSAYVKS1 258
DB 168 NRVTYITQLYKDDPTIFSWQLANEARCNGNPHGLPVK-NCNTDTITKNWDEIATFIHQE 226
QY 259 DPNHMSVGDGEFYIGSTQSGWPNY-----DPSDGVNALLRVKNIDFGTYHLVNTYWG 314
DB 227 DPNHLVSSGIEG--IGLTPPAGVDKNTVYTYTEGTDYEAISALDSIDYNTVHYMPVWG 284
QY 315 --QNAQDKGTQWKDHIANAAGKPTILEEFGWQT-----PDRDSV1 QTWQTQVTRNG 365
DB 285 LKDYAKQGTWTKAHADYDKFKNKPVTVEENGLSTADNPVLEQRDP1YTWMMNEVLAND 344
QY 366 BAGNH--LWMLAGNVNQCQPNYVDG----- 389
DB 345 LAGNH--LWMLAGNVNQCQPNYVDG-----VCGEDYDGLGILLEDEITAVIDPFTKLYANQCNELDTISIVIT 400
DB 422 LAGNH--LWMLAGNVNQCQPNYVDG-----TSPSPSSSP----- 422
DB 460 VVVYVEEGQPKYGTCTGKCKCAHCTRCBGEYVQCQRPITEPPYRGATSPVEGYVL 460
QY 423 -----SSSFPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 470
```

```
DB 461 PGAKSTSKNNNTTKTTTTSKSEPTSSSSDECF5IALGFPCCSDNTVTVVSDND--- 517
QY 471 GGNQIKPGIQLQVNTGSSVD-----LSTVTYVYWFTRDGGSSTLVYNCIMAA 517
DB 518 GDMGVENG-EMGIGIGT1VDNDSCFAKSLGYSCCSDVVY---TDNDGNMGVENGEW-- 571
QY 518 MCGG 521
DB 572 --CG 573

RESULT 14
O24820
ID O24820 PRELIMINARY: PRT: 1000 AA.
AC O24820;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
RT thermophilic anaerobe strain NA10.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB008029; BAA22939.1; -.
DR HSP: O06851; INBC.
DR InterPro: IPR001589; Actbind_actnin.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; GH_5.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00942; CBM_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLYDRLASE10.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 14.8%; Score 609.5; DB 2; Length 1000;
Best Local Similarity 26.8%; Pred. No. 5e-23;
Matches 180; Conservative 116; Mismatches 209; Indels 167; Gaps 31;

QY 156 LTNDWKEFGMDQY--DKWYGI,-----PYHDNFYTDPRTOOAYKNVNV 196
DB 103 VYNEAIDEGSGDYRRSNWNYICGPEYIEKAFIWAHEADPNKLFYNDYNTENSOKR--O 160
QY 197 HLLNRVNST--TGVYTKNDPTIFAWELANPRCVSGTLPTSGCTCTQATI-VNW--VDOM 251
DB 161 FIYNLKSLEKGV-----PIHGVGLQCHINDWPSISEI 195
QY 252 SAYVK--SIDPN---HMVSGDEGYIGSTQSGWPNYNDPSCVDNNALLRVKNIDFGTY 306
DB 196 ENTIKLFSIPGIEIHTEL-DMSFY---QWGSSTNYAVPPRELLIKOARYKEL-FDLF 250
QY 307 HLYPN-----YWGONADWGTOWIKDHTANAAAIK---PTILEEFGWQTPDRDSVYQW 357
DB 251 KKKYKVVNTVTFWGLKDDY--SWLSKN-----YCKDYPLLFDE----- 287
QY 358 TOTVRTNGEAGNFWMLAGNVNQCQPNYVDGPNVYPS---STATVLASEALAISTGSP 414
DB 288 -----NYMSKYAFNSLID-----PSIVPTTTLTPPEIOLTLTPT 324
QY 415 P-----PSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 468
DB 325 PTASVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 384
```

```
QY 469 APCDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSFLVYVNCDAAMGCGNIRASFG 528
DB 385 NSTTNTIRPLKVVNTGSSIDLSTVRYWFTVDGDKAQSASV-DMAQIGASNTVTKFV 443
QY 529 SVNPATPTADTYLQLSFTCGT--LAAGSTGEIQNRVKNKSDMSNFDETNDYSY-GTNTAF 585
DB 444 KLSSSVSGADYYLEIGFKSGAGQLPGKDTGETQIRFNKSDWSNYNOGNDWSWIOSMTSY 503
QY 586 ODWTKVTVVNGRLVNGTEPSC-----TSPSTPSPSPSPSPSPSPSPSPSPSPSPSPSP 639
DB 504 GNMKVATIDGVLVWQDEPTGATAPIATPTPTAPTATPTPTPTPTPTPTPTPTPTPTPT 563
QY 640 PPSPTPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 682
DB 564 PTITAPTAPTAPTSTPAYLDNDNDLVSCNKIVKDG---RPWLTGVNW-FCY--- 616
QY 683 VTVNTGSRATSGVTWVMSFGNQTVTWNTALTQSGASVTA-TNLSYNNVIOQSQSTT 741
DB 617 ----NTGTVFVFG---VWSCNLKSTLAEIANRGENLLRPVISAELILNMSOGIYKPKPIN 669
QY 742 FGFNGSYSCNT 753
DB 670 YVYNPELECKNS 681

RESULT 15
O52374
ID O52374 PRELIMINARY; PRT: 1779 AA.
AC O52374;
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN XYNC.
OS Caldicellulosiruptor sp. R69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Syntrophomonadaceae; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RT69B.1;
RA Morris D.B., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT R69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036924; AAB95326.1; -.
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR005084; CBD_6.
DR InterPro: IPR003305; CBM_GenC.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00942; CBM_3; 3.
DR Pfam: PF02018; CBM_4; 9; 2.
DR Pfam: PF03422; CBM_6; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLHYDRLASE10.
DR PRODOM: PD001947; CBD_3; 3.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase; Xylan degradation.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;
```

Query Match 14.8%; Score 609; DB 2; Length 1779;
Best Local Similarity 25.9%; Pred. No. 9.9e-23;
Matches 189; Conservative 116; Mismatches 213; Indels 212; Gaps 28;

```
QY 25 TLGSITMPSATAAPAG--FVTASGQGVNLGLPYRYGGTNNYYLSYQSHADVDVLAQAQ 82
DB 702 TVSATPTPTAPTASGAGSYWTPSES-----YGALKVWYVNGNMSSTTNVLNPKIK 751
QY 83 AMNLSV-----IRTWGFFID-----IGS-----LDGSVPTIDGNKNG--FYFQ 117
DB 752 IENVGTTAVDLNRVKNRYWYTIIDGAAOSVSVASSINPAYIDRVVVKLGANAGGADYYVE 811
```

```
QY 118 YWDPSS-----TCAPAYNDGPTUGLOLDYALASAAAHLGRVIVVLTN 158
DB 812 VFKSGAGVLAACQSTKEIRLSIQKSSGSYN-----OSNDYSVRSANSY-----IFN 858
QY 159 DMKEFGGMDQYDKWYGLPYHDNFYTDPTQOQAYKKNVNHLLNRVNSITGVYTKNDPTIFA 218
DB 859 E-KVTCYIDDLVW-----GREPGRNAQIKVW--YANGNLSSTPNVL---NPKI-- 901
QY 219 WELANPRCVGSGTLLTS-----CTCQOATIVNVWDOMSAYVKSIDPNH-----M 263
DB 902 -KIEN-----VGTAVDLNRVKNRYWYTIIDGATQSVS-----ASSINPAYIDRVV 947
QY 264 VSVG-----DEGFYIG-----STGSGMPYNDPSPGVNNALLRV 298
DB 948 VKLGANAGGADYYVEVGFKSGAGVLAAGOSTKEIRLSIQKSSGSYNOSNDYSVRSANSY 1007
QY 299 KNIDFGTYHLYPNYWGQNDMGWTQWIKDHITANAAALCKPTILEEFGWQTPDRDSVYQWT 358
DB 1008 ENEKVTGYIDDLVWCKEPSRGT-----KP----- 1032
QY 359 QVVRTNGEAGWPFMWLAGWVNCOPYNYDGFNVVYSSSTATVLAASEALAISTGTSPHPSIP 418
DB 1033 -----AGEVTPAPTPT-----STPTPTPTTPTAPTAPTAPTAPTPTPTPTPTPTPT 1072
QY 419 SSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 477
DB 1073 TPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1132
QY 478 GLQLVNTGSSVDLSTVTVRYWFTRDGSSFLVYVNCDAAMGCGNIRASFGSYNPATPTA 537
DB 1133 WFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDMAQIGASNVTFFVKLTSGVSGA 1191
QY 538 DTYLQLSFTTGGT--LAAGSTGEIQNRVKNKSDMSNFDETNDYSY-----YCTNTAFQDW 588
DB 1192 DYILEYGFSSGAGQLPGKDTGDIQVRFNKNOWNSNTNOADDWSWMSMTNYGEN----- 1245
QY 589 TKTVTVVNGRLVNGTEPSCPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 648
DB 1246 AKVTLYVDGVLVWQDEPTGATAPIATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1305
QY 649 SPSPSPSPSV 658
DB 1306 TPTPTPTKPV 1315
```

Search completed: November 13, 2002, 11:54:07
Job time : 75.6164 secs

GenCore version 5.1.3
Copyright (C) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:49:46 ; Search time 23.5388 Seconds
(without alignments)
952.482 Million cell updates/sec

Title: US-09-917-378-1
Perfect score: 4122
Sequence: 1 MGLRRPARAFVATAAGTAV.....GFNGSYSGTNTAPTLTCTAS 762

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	813	19.7	680	4	US-09-134-078-25
2	812	19.7	663	4	US-09-134-078-61
3	674.5	16.4	521	1	US-08-276-213-3
4	644	15.6	377	1	US-08-525-697-2
5	624	15.1	1751	4	US-09-136-574A-44
6	589.5	14.3	1426	4	US-09-136-574A-43
7	488	11.8	551	2	US-09-033-537A-1
8	469.5	11.4	616	4	US-09-136-574A-47
9	438	10.6	700	2	US-07-862-588B-2
10	415.5	10.1	493	4	US-09-198-956-10
11	415.5	10.1	493	4	US-09-198-955A-12
12	415.5	10.1	493	4	US-09-694-531-12
13	415.5	10.1	493	4	US-09-670-141-10
14	366.5	8.9	167	5	PCT-US95-13813-9
15	279	6.8	381	4	US-09-216-295-22
16	246.5	6.0	108	2	US-08-585-585A-2
17	246.5	6.0	108	2	US-08-249-037C-2
18	246.5	6.0	108	2	US-08-788-622B-2
19	246.5	6.0	108	3	US-08-788-621B-2
20	241.5	5.9	104	2	US-08-585-585A-3
21	241.5	5.9	104	2	US-08-249-037C-3
22	241.5	5.9	104	2	US-08-788-622B-3
23	241.5	5.9	104	3	US-08-788-621B-3
24	230.5	5.6	106	2	US-08-585-585A-4
25	230.5	5.6	106	2	US-08-249-037C-4
26	230.5	5.6	106	2	US-08-788-622B-4
27	230.5	5.6	106	3	US-08-788-621B-4

28	228	5.5	382	4	US-09-277-716-22	Sequence 22, Appl
29	228	5.5	382	4	US-09-609-161B-22	Sequence 22, Appl
30	226	5.5	162	1	US-08-048-164A-2	Sequence 2, Appl
31	226	5.5	162	1	US-08-460-462-2	Sequence 2, Appl
32	226	5.5	162	1	US-08-460-457-2	Sequence 2, Appl
33	226	5.5	162	1	US-08-460-458-2	Sequence 2, Appl
34	226	5.5	162	2	US-08-460-455-2	Sequence 2, Appl
35	226	5.5	162	2	US-08-330-394A-2	Sequence 2, Appl
36	226	5.5	163	3	US-09-006-636-7	Sequence 7, Appl
37	226	5.5	163	4	US-09-006-632-7	Sequence 7, Appl
38	226	5.5	163	4	US-09-325-274-7	Sequence 22, Appl
39	224	5.4	156	2	US-08-330-394A-22	Sequence 22, Appl
40	222	5.4	490	4	US-09-109-841-2	Sequence 4463, Ap
41	218.5	5.3	2137	4	US-09-134-001C-4463	Sequence 29, Appl
42	218	5.3	154	2	US-08-330-394A-29	Sequence 8, Appl
43	214.5	5.2	631	3	US-08-814-052-8	Sequence 8, Appl
44	214.5	5.2	631	3	US-08-812-829-8	Sequence 1, Appl
45	212.5	5.2	109	2	US-08-585-585A-1	

ALIGNMENTS

RESULT 1
US-09-134-078-25
; Sequence 25, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 680 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-134-078-25

Query Match 19.7% Score 813; DB 4; Length 680;
Best Local Similarity 41.5%; Pred. No. 5.3e-48;
Matches 173; Conservative 65; Mismatches 135; Indels 44; Gaps 13;

QY 19 AVAAATLGSITMPSATAAPAGFVTASGGQFVLNGLPYRGCTNNYVLSYQSHADVDVL 78

Db 12 SVSAEFLLLIVLSFVLFASDEFVKVENGKFAENGKFRFGISNNYMHYKSNMGIDSVL 71
Qy 79 AKAQANLNSVIRTWGFDIGSLDGSVPTIDGNKNGFYFOYWDPTCAPAYNDG----PTG 134
Db 72 ESARDMGIKVLRWGR-----LDGESYCRDKN-----TYMHPEPGVFGVPEGTSNAQSG 120
Qy 135 LQCLDYAIAASAAHGLURVIVLTNDWKEFGGMDQYDKWYGLPYHDFYTDPRTOQAYKNN 194
Db 121 FERLDYTVAKAKELGKLVILVNNWDDFGGMNQYVRWFGGTHDDFYRDEKIKKEYKKY 180
Qy 195 VNHLLNRVNSITGVYTKNDPTIFAWELANEPKRCVSGTLPTSGTCTQATIVNWDOMSAY 254
Db 181 VSLFVNHVNTYGVVREPTIMAWELANEPKRC-----ETDKSGN-----TLVEWVKEMSSY 232
Qy 255 VASIDPNHMSVSGDECF-----YIGSTCGSG---WPYNDPSDGVNALLRVKNIIDFGTYHL 308
Db 233 IKSLDPNHLVAVGDEGFFSNYEGFKPYGGEAEWAYNGWS-GVDWKLLSLETVDGTFHL 291
Qy 309 YPNYWGON----ADWCTOWIKOHIANAAIGKPTILEEFG--WQTP-DRDSVYQVWTQV 361
Db 292 YPSHMCVSPENYAQMCAKWIEDHIIKIAKEIGKPVVLEEYGIKPSAPVNRRTAIYRLWDLV 351
Qy 362 KTNCEAGHFWMLAGNVNQP-----YPNYDGFNVYPSSTATVLAASE-ALAISTG 411
Db 352 YDLGGDCAMFWMLAGIGESDRDERGYYPDYDGFIRVNDSDSPEAEIIRYAKLFNTG 408

RESULT 2
US-09-134-078-61
: Sequence 61, Application US/09134078
: Patent No. 6368844
: GENERAL INFORMATION:
: APPLICANT: Bylina, Edward J.
: TITLE OF INVENTION: GLYCOSIDASE ENZYMES
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gray Cary Ware & Freidenrich LLP
: STREET: 4365 Executive Drive, Suite 1500
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/134,078
: FILING DATE: 13-AUG-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/949,026
: FILING DATE: 10-OCT-1997
: APPLICATION NUMBER: 60/056,916
: FILING DATE: 06-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 858/677-1456
: TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 61:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 663 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: "PAGE" TYPE: Internal
NS-0 134-078-61

Query Match 19.7%; Score 812; DB 4; Length 663;
Best Local Similarity 42.8%; Pred. No. 6e-48;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13:
Qy 41 FVTASGGQFVNLGCLPYRYGCTNNYLSYOSHADVDDVLAKAQAANLNSVIRTWGFDIGSL 100
Db 17 FVKVENGKFAENGKFRFGISNNYMHYKSNMGIDSVLESARDMGIKVLRWGR-----L 71
Qy 101 DGSVPTIDGNKNGFYFOYWDPTCAPAYNDG---PTGLOGLDYAIASAAHGLKVLVVL 156
Db 72 DGESYCRDKN-----TYMHPEPGVFGVPEGISNAOSGFERLDYTVAKAKELGKIKVIL 125
Qy 157 TNDWKEFGGMDQYDKWYGLPYHDFYTDPRTOQAYKNNVNHLLNRVNSITGVYTKNDPT 216
Db 126 VNNWDDFGGMNQYVRWFGGTHDDFYRDEKIKKEYKKYVSLVNHVNTYGVVREPT 185
Qy 217 FAWELANEPKRCVSGTLPTSGTCTQATIVNWDOMSAYKVSIDPNHMSVSGDEGF----Y 273
Db 186 MAMELANEPRC-----ETDKSGN-----TLVEWVKEMSSYIKSLDPNHLVAVGDEGFFSNYE 237
Qy 274 GSTQSGS---WPYNDPSDGVNALLRVKNIIDFGTYHILYPNYWGON----ADWCTOWIKD 326
Db 238 GFAPYGEAEWAYNGWS-GVDWKLLSLETVDGTFHILYPSHMCVSPENYAQMCAKWIED 296
Qy 327 HTANAAIGKPTILEEFG--WQTP-DRDSVYQVWTQVTRTNGEAGMFWMLAGNVNQP- 382
Db 297 HIIKIAKEIGKPVVLEEYGIKPSAPVNRRTAIYRLWDLVYDLGGDCAMFWMLAGIGESDR 356
Qy 383 -----YPNYDGFNVYPSSTATVLAASE-ALAISTG 411
Db 357 DERGYYPDYDGFIRVNDSDSPEAEIIRYAKLFNTG 391

RESULT 3
US-08-276-213-3
: Sequence 3, Application US/08276213
: Patent No. 5536655
: GENERAL INFORMATION:
: APPLICANT: Thomas, Steven
: APPLICANT: Laymon, Robert
: APPLICANT: Himmel, Michael
: TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: National Renewable Energy Laboratory
: STREET: 1617 Cole Boulevard
: CITY: Golden
: STATE: CO
: COUNTRY: USA
: ZIP: 80401-3393
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/276,213
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: O'Connor, Edna
: REGISTRATION NUMBER: 29,252
: REFERENCE/DOCKET NUMBER: NREL IR# 94-08
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303)231-1000
: TELEFAX: (303)231-1098
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 521 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein

;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 16.4% Score 674.5; DB 1; Length 521;
Best Local Similarity 27.2% Pred. No. 1.2e-38;
Matches 210; Conservative 88; Mismatches 177; Indels 297; Gaps 32;

```
QY 37 APAGFVTSAGGQV--LNGLPYRYGCTN-----NYLSYOSHIADVDVLAKAQAMNLSV 88
DB 1 AGGYWHTSCREJLDANNVPVRIAGINWCFETCTNVHGLWSRDYKSLMDQIKSLCYNT 60
QY 89 IRTWGFIDICSLDSVPTIDGNKNGFYFOYWDPSDTCAPAYNDGPTGLOGLDYAIASAAAH 148
DB 61 IRL-PSDDILKPTNP-----NSINFYQMNQD-----LQGLTSLQVMDKIVAYAGQI 107
QY 149 GURVIVVLTDNMKEFCGMDQYDKWGLPYHDNFYDPTQQAYKNVHLLNRVNSITGV 208
DB 108 GLRII--LDRHRPDCSG--QSALWY-----TSSVSEATWISDL-----QALAQ 146
QY 209 TYKNDPTIFAWELANEP---CVGSGTLPTSGCTOATIVNW---VDQMSAYVKSIDPNH 262
DB 147 RYKGNPTVVGFDLHNEPHDPACWCGCD-PS-----IDWRLAAERAGNAVLSVNPNL 196
QY 263 MVSU-----GDEIFYIGSTGSG-WPYNDPSDGVDDNALLRVKN-----IDFGTYHL 308
DB 197 LIFVEGVQSYNGDSYHWGNLQAGQY-----VVLNVPNRLVYSAHDYAT-SV 244
QY 309 YPNWYQONAD-----NCTOW---IKDHIANAAAIKPTILEEFGWOTPPDRDSVYQW 357
DB 245 YPOTWFSDPTEFPNNMPCINWKNNGYLFNQIA-----PVMLGEGF--TTLQSTTDQW 295
QY 358 TOTVRTNGEAGNFNWLACNVNCPYNYDGFNVYPSSTATVLAASEALAIAGTSPPPS 417
DB 296 LKTL-----299
QY 418 PSSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 477
DB 300 -----VQY-----LRP 305
QY 478 GLQVNTGSSVDLSTVTVYVWFTRDGGSTLVYNCDAAMCGCNTIRASFGSVNPAITPA 537
DB 306 -----TAQY-----CADSPQWTFWSNPDSPS---325
QY 538 DTYLQLSETGGTLAAGSGTGEIONRVNYSKDSWNEFDETDNYSYCTNTAFQDWTKVTVYVNG 597
DB 326 -----GDTGGIL-----KDDMTQVDTVKO-----GYLAPIKSIF---355
QY 598 RLVMCTEPTSGTSPPTPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 651
DB 356 -----DPVGASASPSQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 409
QY 652 PPSPPS-VSSSGVGCRATVYVNSDWCSGTATVTVNTGSRATSGTGVTVAMSGGNTVTN 710
DB 410 ASPTSPPTAASGARCSTASYOVNSDMGNGFTVTVAVTNSGVSATKTWTVSWTFGGNQTITN 469
QY 711 YWNTALTQSGASVTATNLSYNNVIOPGQSTTFEFGNSYSYGTNTAPTTLCTAS 762
DB 470 SNNAAVTQNGQSVTARNMSYNNVIOPGQNTTFEFGQASVTSNNAAPTIVACAAS 521
```

RESULT 4
US-08-525-697-2
; Sequence 2, Application US/08525697
; Patent No. 5795764
; GENERAL INFORMATION:

; APPLICANT: Christgau, Stephan
; APPLICANT: Andersen, Lene N
; APPLICANT: Kauppinen, Sakari
; APPLICANT: Heldt-Hansen, Hans P
; APPLICANT: Dalboege, Henrik
; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY

;
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5795764o No. 5795764disk of No. 5795764th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,697
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4004,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0486/93
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-525-697-2

Query Match 15.6% Score 644; DB 1; Length 377;
Best Local Similarity 35.2% Pred. No. 1e-36;

Matches 138; Conservative 67; Mismatches 147; Indels 40; Gaps 10;

```
QY 24 ATLGST---MPSATAAPAGFVTASGGQVNLGLPYRYGCTNNYLSY-QSHADVDVL 78
DB 11 ASLGVATALPRTPHNNAATTAFTPSTGLHFTIDCKTGYFAGTNSYWGFLTNDDVDLVM 70
QY 79 AKQAAMNLSVIRTWGFDIGS--LDGSVPTIDGNKNGFYFOYWDPSDTCAPAYNDGPTGLQ 136
DB 71 SOLAASDLKILRVMGFNDVNTKPTDGTV-----WYOLHANGTSTINTGADGLQ 118
QY 137 GLDYATASAAAHGLRVIVLTDNMKEFGMDQYDKWGLPYHDNFYDPTQQAYKNV 196
DB 119 RLDYVVTSAEKYGVKLLINFNWNTDYGGMQAYVTAYGAAATDFTYNTAIAQAYKNYIK 178
QY 197 HLLNRVNSITGVYKNDPTIFAWELANEPKVCVSGTLPTSGTCTQATIVNVDQMSAYVK 256
DB 179 AVVSR-----YSSAAAIKAWELANEPKVCVSGTLPTSGTCTQATIVNVDQMSAYVK 221
QY 257 SIDPNHMSVGDGFGFYIGSTGCGSWPNDSQVDNALLRVKNIDFGCTYHLYPNYHQN 316
DB 222 SLDKSLVITIGDEFGLDVDSGSGPYT-YGEGLNFTKNLGISTIDPDTLHLYPDMSGTS 280
QY 317 ADMGTQWIKDHIANAAGIKPTILEEFGWOTPPDRDSVYQWOTQTV-RTNCEAGNFWMLA 375
DB 281 YDWNGCMITAHAAACKAVGKPCLLLEEG-VTSNHCAPESWQOTAGNATGISDLYHGYG 339
QY 376 GNVNGQPYNYDGFNVYPSSTATVLAASEALA 407
DB 340 TTFWSMGQSPN-DGNTFEYNTSDFTCCLVTDHVA 370
```

RESULT 5
US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:


```

: APPLICANT: Farrington, Graham K.
: Anderson, Paige
: Gibbs, Moreland
: Bergquist, Peter
: Daniels, Roy
: Morgan, Hugh W.
: Williams, Diane P.
: TITLE OF INVENTION: Compositions and Methods for
: Treating Cellulose Containing
: Cellulase Enzyme Compositions
: NUMBER OF SEQUENCES: 49
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Howson and Howson
: STREET: Spring House Corporate Center, P.O. Box 457
: CITY: Spring House
: STATE: PA
: COUNTRY: USA
: ZIP: 19477
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/136,574A
: FILING DATE: 19-Aug-1998
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/932,571
: FILING DATE: September 19, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Bak, Mary E.
: REGISTRATION NUMBER: 31,215
: REFERENCE/DOCKET NUMBER: 1997US001/CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-540-9200
: TELEFAX: 215-540-5818
: TELEX: <Unknown>
: INFORMATION FOR SEQ ID NO: 44:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1751 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 44:
: US-09-136-574A-44
:
Query Match          15.1%; Score 624; DB 4; Length 1751;
Best Local Similarity 25.6%; Pred. No. 1.8e-34;
Matches 234; Conservative 140; Mismatches 278; Indels 262; Gaps 46;

QY      7 PARAFVA-TAAGTAVAAAAATLGSITMPSA-----TAAPAGEVFTASG---- 46
       || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     183 PASSVVAETAASLAASVAAWVKIKERNSSQAAYSLQHAKDLFEFADTTSDAGYTAATGCTYS 242
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
QY     47 GQFVLN--GLPYRYGTNNYLKSYLSYSHADVDDVLAKAQAMNLNVIRTWGTFIDIGSL--- 100
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db    243 GGFIDDLGWAHVWLYIATND--SSYLTKA--EELMSEYANGNTWTQCWDVDVRYGTLIML 298
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
QY    101 -----DCGSPTIDGNKNNGFYQYW-DPSTCAPAYNDGPTQLGCLDYAIASAAAHGL 150
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db   299 AKITGKEYLKGV-----ERN---LDHWTDRITYTPKGMAVLTGWGSLRYATTAA----- 345
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
QY    151 RVIVWLTDNHKEFGCGMDYDKW-----YCLPHYDNFYDPTQTQA 190
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db   346 -FLACVVYADW---SGCDNSNKKTLYLNFAKSQIDYALGSTGRSFVGFGTYNPQPHIRNA 401
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
QY    191 YKNWVN-----HLLNRVNSITG-----VTYKNDPTIFAWELANEPRC-VGSGTLPT 235
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db   402 HSSWANSKMKIPEYHRHL--YGALVGGPCGSDSYNDIDTY---VQNEVACDYNAGIVGA 456
       | | : | | | | | | | | | | | | | | | | | | | | | | | |
QY   236 SGTCCTQATINWVDQMSAYVKSIDPNHMVSVGDGEFY---IGSTQG-----S 279
       | | : | | | | | | | | | | | | | | | | | | | | | | | |

```

```

457 LAKMYQLYGGEPIDDF-----KAIE-----TPTNDEIIFVESKEFGNSQCPNYYTEVYSIYNKNT
280 GWP-----YNDPSDGVDDNNALLRVKNIIDFGTYHLY-----PNYWGONAD----318
459 GWPFRVTDKLSFKYFIDLTETELIQAGYSPPDVVKVD---TYIIEGCKISGPIYWDKNNKIYYV566
319 ----WGOWIK-----DHIAANAATIGKPTILIEEFCWOTPPDRDSYVYQWTOTVTRTNGEAGNN370
567 LVDFSGTKIYPGGVEHKKQAQF--KISVPOGYPWD--PTNDPSYKGLTQJLEKN-----617
371 FWMLAGVNGCGPPNYDGFNVYPPSSSTATVLAESALATGCTSPPPSPSSSPSSSPSSPSPSP430
618 -----KYIAAYDNNNLVW-----GLEPGAATSTP-----APTSTPTPTP651
431 SPSASPSASPSASPSASPSASPSASPSVSG-GVKVOYKKNDSAPGDNOIKPGLQLVNTGSSSV489
652 TPTPTVATP--TPTPTPTPGSPCTGSGVKVLYKKNETSASTGSIHPWFKIYNGJSSSV709
490 DLSTVTYVYWFTRDGGSTLIVYNCDMAAMCCGNIASFSGSVNPATPTADTYLQLSFTGCT549
710 DLSRVKIRYWTYVDGDKPOSAN-CDMAQICASNVTNFYKLSGCVSGADYYLEVGFSSKA768
550 --LAAGSTGEIQNRVKNKSWSNFDENYDS-----YCTNFAQDWTKVTVYVNGKLV600
769 GQLQPGKDTGDIQYRFKNNKDNWSYNOADDWSLQSMNTNYGEN-----AKVTLYVDJVLV822
601 WGTPEPCTSPSPPTSPSPPTSPSPPTSPSPSSPSPSPSPSPTSPSPSP-----SP654
823 WQCEPGGATPAPTSTATPTPTPTATPTPTP--TPTPTPTVSATPTPTAPTASPVGGSYWTPT880
655 SPVSYSVGCGCRATPYVYNSDWSG--FTATVTVTNTGS-----RAT693
881 SES-----YCALKVMYANGNLSSPTNVLNPKIKIENVGTTAVDLRSVKVRYWYTIIDEAT945
594 SCWTVANS-----FGNOTVYNTVNTALTOSGASVATNTLNSYNNV---IOPKO718
936 QSVSVASSINPAYIDVKEFVKLGANAGGADYYVEIGFKSGAGVLAAGOSTKEIRLSIQKG-994
739 STTFGNGSYSGTN752
995 -----SGSNQSN1002

RESULT 6
US-09-136-574A-43
: Sequence 43, Application US/09136574A
: Patent No. 6294366
: GENERAL INFORMATION:
: APPLICANT: Farrington, Graham K.
: Anderson, Paige
: Gibbs, Moreland
: Bergquist, Peter
: Daniels, Roy
: Morgan, Hugh W.
: Williams, Diane P.
: TITLE OF INVENTION: Compositions and Methods for
: Treating Cellulose Containing Fabrics Using T
: Cellulase Enzyme Compositions
: NUMBER OF SEQUENCES: 49
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Howson and Howson
: STREET: Spring House Corporate Center, P.O. Box 457
: CITY: Spring House
: STATE: PA
: COUNTRY: USA
: ZIP: 19477
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/136,574A
: FILING DATE: 19-Aug-1998

```

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6294366e
SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43

Query Match 14.3%; Score 589.5; DB 4; Length 1426;
Best Local Similarity 24.6%; Pred. No. 3.3e-32;
Matches 199; Conservative 138; Mismatches 251; Indels 221; Gaps 33:

Qy 84 MHLVSIVRTWGFIDF---GSLGDSVPTI-DGKNKGFYFQWDPSP-----T 123
Db 19 MSISLVGMSYFPVKTEAAPDWSISLWESYKNDKFGVAPARCLSNDDTKQMWLKIKN 78
Qy 124 GAPAYND-----GPTGLOGLOVAISAAAH-----GLRVIVVL-----TNMKKE 162
Db 79 SITAENEMKPESSLLAGQTS-TGLSYRFSTADTFVFANFTNNIGIRGHTLVVHNNQTPDW-- 135
Qy 163 FCGMDQYDKWGLPYHDFNYDPTQQAYKNWVHLLARVNS-LTGVIYKNDPTIFAWEL 221
Db 136 -----FFRDSQQMLSK---DALLARLQYIYDVVGRYKGYVYAMDV 174
Qy 222 ANEPRCVGSGTLPTSGCTQATIVNWVDMSAYKSIDPNHMSVSGDECFYIG----- 274
Db 175 VNEAIDESQPCYRRSTWYQICGPEY IEKAFIWAHEADPNAKLFYNDYNTLSTKRDFIY 234
Qy 275 -----STQSGWP-----YNDPSGDVDNNAL-----LRVKNIDFCTYHLY 309
Db 235 NMVKNLKSQVPIHIGIMQSHINWMPVSSEIENSKLFSSIPGIEIHTELDMSLYN-- 292
Qy 310 PNYWCONADWGTQWIKDHIANAATGKP--TILEPFG-----WQTPDRDSVYQTWQ 359
Db 293 ---YGSNENYSTP-PQDLLQROAQYKIDFTWLRKYKGIIVCTVTEWGLKDDYS----- 341
Qy 360 TVRTNGEAGWFMWLAGNVNGOYPNYDGFNVYYPSPSTATVLASPA---SGTSPPPSPS 419
Db 112 -----WLNSSSKRDWPLLFEDDYSA-KPAYWSVIEAACASA-----SPS 379
Qy 420 SSPSSSP 478
Db 380 PWTATPT 433
Qy 478 VLVHTGSSVVDLSTVTVYVWTFRTDGGSTLVYNCDAAMCGNIRASFGSVNPATPTAD 538
Db 434 PKIVNGGSSVVDLSRVKIRYVTVYDGDQPSAV-CDMAQIAGSNVTFNFKVLSVSGAD 492
Qy 539 TYIOLSTFGGT--LAAGSGTGEIQRNVKNSDMNFDNDS-----YGTNTAFQDWT 589
Db 493 YYLEVGFSSGAGQLQPGKRDAGDIQVFNKNKDNWSYNOADDWSLQSMYDGEN-----A 546
Qy 590 KVTYVNGRLVWGTGCTSP 649
Db 547 KVTLYVDCVLVWQEPGCGATPATATATPTPIPTATVTPPTATPTSPRPTATATPTPT 606
Qy 650 PSPSPSPSVSSSGVCCRTATYVNSD-----WGSGETATVT-----VTNTGS--- 690
Db 607 VSATPTAPTASPVG--GSYWTSPSEYALKVWYANGNLSSTFVNLNPKIKIENWGTTAV 664

Qy 691 -----RATSGWTVAHS-----EGGNOTVTNYWNTALTQSGASVTAT 726
Db 665 DLSRVKRVYWTIDGATOSVSASSINPAYIDVKILGANAGGADYYVEIGFKSGAGVLA 724
Qy 727 NLSYNNV---IQPGQSTTFGFGNGSYSGTN 752
Db 725 GOSTREIRLSIOKG-----SGSYNOSN 746

RESULT 7
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033.537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

Query Match 11.8%; Score 488; DB 2; Length 551;
Best Local Similarity 25.2%; Pred. No. 8.8e-26;
Matches 157; Conservative 86; Mismatches 165; Indels 216; Gaps 25:

Qy 125 APAYNDGPTGLOGLDYA-----IASAAHGLRVIVVLTNDWKEFGMDQYDKWYG-- 174
Db 1 APAPFVGOLKVQGNOLVGQSQOAVQLVGNSSHGL-----OWYGNF 40
Qy 175 -----LPVHDNFYDTPRTQQAYKNWVN-----HLL 199
Db 41 VNKSSLOWNRDNGINVFRAAMYTAEDGYIDPSVKNKVEAQASIDLGLVVIDWHL 100
Qy 200 NRVNSIT-----GVTYKNDPTIFAWELANEPKRCVSGTLPTSGTCTQATIVNW 247

```
Db 101 SCGNPNYKQAKATFOEMATLYGNTPNVI-YEIANEP-----NGNWSWADVKS 149
QY 248 VQMSAYVKSIDPNHMSVSCDE-----CFYIGS----- 275
Db 150 ABEVITAIRAIIDPGVIVGSPWSDIHLAADNPVSHSNVYALHFYSGTHCQFLRDI 209
QY 276 -----TQSGWPNYDPS-DGVNNDALLRVKNIDFGTYHLYPNY--- 312
Db 210 TYAMNKGAAIFVTEWCTSDASGNGGPFQSKWEID---FLNARKLSWVNWSLADKVEYS 266
QY 313 -----W--CONADKCTONIKDHI-----ANAAIGKPTILEEFGWOTPD 350
Db 267 AALMPCASPTGWTDAQLSESG-KWVRQOIROATGGGSGNPTAPAAPTNLS-----ATAGN 321
QY 351 DSVYQVOTVOTVARTNGAG-WNEWMLAGNVNGQPPNYDGFNVYVPSSTATVLASEALAS 409
Db 322 AOVSLTW-----NAVSGATSYTVKRATTSGGPTYN-----VA 353
QY 410 TCTSPPPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 464
Db 354 TGVATSYTNTGTNTGTTTYYVVSASNSAGSSANSQA---SATPASGCASTGNLVQYK 410
QY 465 NNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTDRGGSTLVYNCDMAWCCGNIR 524
Db 411 VGDTSATDNQMKPSPNIRKNGTTPVNLGKLRYYETKD-GTADMSASFDMQOIGASNV 469
QY 525 ASFGSNPATPTADTYLQLSFT--GTLAAGSTGEIQNRVKNKSDMSNFDNYSY-CT 581
Db 470 AAF--ANFTGTSNTDYVELSFSGAGSGSIPAGGQTGDIQLRMKYKTDMSNFNEANDYSYDCA 527
QY 582 NTAFOQWTKVTVVNGRLVWGTGP 605
Db 528 KTAAYADNRVTLHQGTGLWGTTP 551

RESULT 8
US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 616 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 11.4%; Score 469.5; DB 4; Length 616;
Best Local Similarity 34.2%; Pred. No. 1.9e-24;
Matches 108; Conservative 55; Mismatches 112; Indels 41; Gaps 10;

QY 455 VSGGVKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTDRGGSTLVYNC 514
Db 1 MSGGVKVLKKNETSASTGSIIRPFKIVNGSSVDLSRVKIRYWTVDGDKPOSAY-CD 59
QY 515 WAAMCGCNIRASFGSNPATPTADTYLQLSFTGGT--LAAGSTGEIQNRVKNKSDMSNFD 572
Db 60 WAQIGASNVTFNFVKLSSGVSGADYYLEVFGSSGAGLOPGKDTGDIQVRFNKNDMSNYN 119
QY 573 ETNDYS-----YGTNTAFQDWTKVTVVNGRLVWGTGPSPSPSPSPSPSPSPSPSP 675
Db 120 QADDMSWLOSMTNYGEN-----AKVTLVXGVLVWQEPGG---AVTPTSTPTPVSSST 170
QY 626 PTPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 675
Db 171 PPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 230
QY 676 GSGFTATVTVTNGSRATSGMTVAMSGFGNQTVTVNTALQTOSGASVTA-TNLSYNNVI 734
Db 231 -FGF-----NTGTNVFDG---VNSCNLKSALAEIANRGNLLRPISAEILNNKSKI 279
QY 735 OPGOSTTFCGNGSYSG 750
Db 280 YPKPNINYVYVNELEG, 295

RESULT 9
US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862.588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
```

```

: APPLICATION NUMBER: PCT/DK91/00013
: FILING DATE: 18-JAN-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Zelson, Steve T. / Lambiris, Elias J.
: REGISTRATION NUMBER: 30,335 / 33,728
: REFERENCE/DOCKET NUMBER: 30,335 / 33,728
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212 867 0123
: TELEFAX: 212 867 0298
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 700 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-862-588B-2

Query Match          10.68; Score 438; DB 2; Length 700;
Best Local Similarity 25.08; Pred. No. 3.3e-21;
Matches 187; Conservative 109; Mismatches 253; Indels 200; Gaps 34;

QY 5 RPARAFVATAAGTAVAAATLGSIIMPSTATAAPAGFVTASGGOFVLNGLPYRYGGTNNY 64
DB 4 RQRRLRFVS-----AALAVSLTM-TVPMPSVNAASDVT-----FTTQSERAAISPNI 53
QY 65 YLSYQSHADVDVLAKAAMN-----LSVIRTWGFDIGSL--DGSVPTIDGNK 111
DB 54 YGTNODLSGTEWSSRLGGNRLTGYWNENASSAGRDWLHYSDDFLCGNGGVPTDCDK 113
QY 112 NGFYQWIDSTGAPAYNDGPTGLOGLDYATASAAHGLRVIVYLTNDWKEFGMDQYD- 170
DB 114 -----PGAVVTAFHD--KSLENGAYSIVT-----LQMGAYVSRD--KNGPVDSEET 155
QY 171 ----KWYGLPYHDN----FYTDPRTOQAYKN-WYNHLLNRVNSITGVYKNDPT---IFA 218
DB 156 APSPRMDKVEFAKNAPSLOPLDNDGQVYDVEYNFLNR-----YGNASTSTGIKA 207
QY 219 WELANERPCVSGTLPT--SGTCQATIVNVVDOMSAIVKSIDPNHMV---SVGDEGYI 273
DB 208 YSLDNEP-ALWSETHPRTHPEQLOAAELVAKSIDLSKAVKNVDPAEIEFGPALYFGAYL 266
QY 274 GSTOGSGWP-----YNDPSDGV---DNNALLRVKNIDFGYHLYPNYWG---- 314
DB 267 SLQDAPWPSLOGNYSWFIDYILDMKNAHTQNGKRLDVLVD-----HWYPEAQGGGQR 321
QY 315 -----QNADWGTQWIKDHIANAAAI-----GKPT 338
DB 322 IVFGAGNIDTKARVQAPRSLWDPAYQEDSWIGTWFSYLLPLPKLQSSITQYYPGTL 381
QY 339 ILEEF-----GHQTPDRDSVYQTTQTVRTNGEAGNFWMLAGNVN-----GOPY 383
DB 382 AITEFSYGGDNHISGGIATADALGIFQY-----GVYAANYWQTEDNTDYTSAAYKLY 434
QY 384 PNVDGNNVYVPS-----STATVLAS-----EALA--TSPPSPSSSPS- 423
DB 435 RNYDGNKSGGSIKVDAATSDTENSIVYASVTDEENSELHILIVLNKHFDDPINATFOLSG 494
QY 424 -----SSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 460
DB 495 DKTYTSGRVNGFDQGTSDITEQAATINNNQFTYTLPLLSAYHVLK...S...EPVNSDLV 554
QY 461 VOYKNND-SAPGNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSGSTLVYNCDMAAG 519
DB 555 VOYKDGDRNNATDQIKPHFNIONKQSPDLSSLTLRYYFTKD-SSAAHNGWIDWAKLG 613
QY 520 CGNIRASFGSNPATPTADTYLQLSFT--GGTLAAGSGTGEIONRVNKSQWSNFDETNDY 577
DB 614 GSNLIQISFGNHGA--DSDTYAELGFSGGAGSIAEGQSGEIQLRMSKADWSNPFNEANDY 671
QY 578 SY-GTNTAFODTKVTYVYVNGRLVWGTPE 605
DB 672 SFDCAKTAYIDWDRVTLYDQGLVWGTPE 700
```

```

RESULT 10
US-09-198-956-10
: Sequence 10, Application US/09198956
: Patent No. 6165769
: GENERAL INFORMATION:
: APPLICANT: Andersen, Lene N.
: APPLICANT: Schulein, Martin
: APPLICANT: Lange, Niels Erik K.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Schnorr, Kirk
: TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
: FILE REFERENCE: 5377-200-US
: CURRENT APPLICATION NUMBER: US/09/198,956
: CURRENT FILING DATE: 1998-11-24
: EARLIER APPLICATION NUMBER: 1344/97
: EARLIER FILING DATE: 1997-11-24
: EARLIER APPLICATION NUMBER: 60/067,240
: EARLIER FILING DATE: 1997-12-02
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 10
: LENGTH: 493
: TYPE: PRT
: ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match          10.18; Score 415.5; DB 4; Length 493;
Best Local Similarity 27.38; Pred. No. 7.3e-21;
Matches 112; Conservative 71; Mismatches 126; Indels 101; Gaps 14;

QY 247 WVDQMSAYVKSIDPNHMVSVGDEGYIG-----STOGSGWPY-----NDUS 287
DB 135 WVDHNELY-----HSLNV-DKDYDGLDFVCKRDAEYITFSWYVHDGKWSMLMGSSDS 186
QY 288 DGVD-----NNALLRVKNIDFGYHLYPNYVQGNADWGTQWIKDHIANAAAI CKP 337
DB 187 DYNRTITTFHHNWFENLSRVSPFRFCGHIYNNYFNKIIDSG----- 229
QY 338 TILFEGHQTDRDSVYQTTQTVRTNGEAGNFWMLAGN--VNGO---PYPNVDGPNVY 392
DB 230 -INSRMGARIRIENLFENAKDPIVSWYSSSPGYVHVSNNKFNVSRCGSMPTTSTTTYNPP 288
QY 393 YPSSTATV-----LASEALAISTCTSPSPSSPSSPSSPSPSPSPSPSPSPSPSPSPSPSP 447
DB 289 YSISLDNVNVKSLVKONAGVGKIQRRPPTP--TPTSPSP----- 326
QY 448 PSPSSPSPSGVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSS 507
DB 327 ---ANTPVSGNLKVEFYNSNPDDTNSINPOFKVTNTGSSAIDLSKLTRLRYYYTVDCKD 383
QY 508 TLVYNCWAAH-----CGGNIRASFGSVNPNPATPTADTYLQLSFTGGTLAAGSGTGE 558
DB 384 QTFW-CDHAAIIGSGSYNGITSNVKGTFTVKMSSSTNNADTYLEISFTGTTLEPCAHV-Q 441
QY 559 IONRVNKSQWSNFDETNDYSYGTNTAFODTKVTYVYVNGRLVWGTPESGT 608
DB 442 IQGRFAKNDNSYNTOSNDISFKRSQPFVMDQVATYLVNGVLVWKGPEGGS 491
```

```

RESULT 11
US-09-198-955A-12
: Sequence 12, Application US/09198955A
: Patent No. 6187580
: GENERAL INFORMATION:
: APPLICANT: Andersen, Lene N.
: APPLICANT: Schulein, Martin
: APPLICANT: Lange, Niels E.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Moller, Soren
: APPLICANT: Glad, Sanne O. S.
: APPLICANT: Kauppinen, Markus S.
```

APPLICANT: Schnorr, Kirk
APPLICANT: Kongsbak, Lars
TITLE OF INVENTION: No. 6187580el Pectate Lyases
FILE REFERENCE: 5378-200-US
CURRENT APPLICATION NUMBER: US/09/198, 955A
CURRENT FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: 1343/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 1344/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/067, 249
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 60/067, 240
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 09/073, 684
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 09/184, 217
PRIOR FILING DATE: 1998-11-02
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 493
TYPE: PRT
ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 10.1%; Score 415.5; DB 4; Length 493;
Best Local Similarity 27.3%; Pred. No. 7.3e-21;
Matches 112; Conservative 71; Mismatches 126; Indels 101; Gaps 14;

QY 247 WVDQMSAYVKSIDPNHVMVSGDEGYIG-----STGSGWPY-----NDPS 287
DB 135 WVDHNELY-----HSLNV-DKDYDGLFDVVKRDAEYITFSWNYVHDGKMSMLMGSSDS 186
QY 288 DGVN-----NNALLRVKNIDFCTYHLYPNYWGONADWGQWIKDHIANAAIGKP 337
DB 187 DYNRTITTFHHNFENLSRVSPFRGEGHLYNNFKIIDSG-----229
QY 338 TILFEFCWOTPDSDSVYQVOTQTVRTNGEAGWFMMLAGN--VNGQ---PYPNYDGFNVY 392
DB 230 -INSRMCARIRIENLLENFENAKDPIVSWSSPGVHWYNNKFNKNSRGMPTTSTTTNNPP 288
QY 393 YPSTATV-----LASEALAISTGTSTPPSPSSPSSPSSPSSPSSPSSPSSPSSPSS 447
DB 289 YSYSLDNDVNVKSVIKONAGVGKIQRRPPTP--TPTSPPS-----326
QY 448 PSPSSSPVSGVGVQYKKNDSAPCDNOIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSS 507
DB 327 ---ANTPVSNLKVFEFYNNSPDTTNSINPQFKVTNTGSSAIDLKSLTLRYYYTVDGQKD 383
QY 508 TLVYNCDAAM-----CGGNIRASFGSVNPATPTADTYLQLSFTGTLAAGSGTGE 558
DB 384 QTFW-CDHAAIIGSNGSYNGITSNVKCTFVKMSSTNNADTYLEISFTGTLPEGAHV-Q 441
QY 559 IQRVKNKSDWSNFDNDETNDYSGTNTAFQDWTKVTYVYNGRLVWGTEPSGT 608
DB 442 IQGRFAKNDSNVTQSDNYSFKRSQFVENDQVTAYLNGVLVWGKEPGGS 491

RESULT 12
US-09-694-531-12
Sequence 12, Application US/09694531
Patent No. 6368843
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schuelein, Martin
APPLICANT: Lange, Niels E.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Molier, Soren
APPLICANT: Glad, Sanne O. S.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Schnorr, Kirk
APPLICANT: Kongsbak, Lars

TITLE OF INVENTION: No. 6368843el Pectate Lyases
FILE REFERENCE: 5378-200-US
CURRENT APPLICATION NUMBER: US/09/694, 531
CURRENT FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 09/198, 955
PRIOR FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: 1343/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 1344/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/067, 249
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 60/067, 240
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 09/073, 684
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 09/184, 217
PRIOR FILING DATE: 1998-11-02
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 493
TYPE: PRT
ORGANISM: Clostridium thermocellum
US-09-694-531-12

Query Match 10.1%; Score 415.5; DB 4; Length 493;
Best Local Similarity 27.3%; Pred. No. 7.3e-21;
Matches 112; Conservative 71; Mismatches 126; Indels 101; Gaps 14;

QY 247 WVDQMSAYVKSIDPNHVMVSGDEGYIG-----STGSGWPY-----NDPS 287
DB 135 WVDHNELY-----HSLNV-DKDYDGLFDVVKRDAEYITFSWNYVHDGKMSMLMGSSDS 186
QY 288 DGVN-----NNALLRVKNIDFCTYHLYPNYWGONADWGQWIKDHIANAAIGKP 337
DB 187 DYNRTITTFHHNFENLSRVSPFRGEGHLYNNFKIIDSG-----229
QY 338 TILFEFCWOTPDSDSVYQVOTQTVRTNGEAGWFMMLAGN--VNGQ---PYPNYDGFNVY 392
DB 230 -INSRMCARIRIENLLENFENAKDPIVSWSSPGVHWYNNKFNKNSRGMPTTSTTTNNPP 288
QY 393 YPSTATV-----LASEALAISTGTSTPPSPSSPSSPSSPSSPSSPSSPSSPSSPSS 447
DB 289 YSYSLDNDVNVKSVIKONAGVGKIQRRPPTP--TPTSPPS-----326
QY 448 PSPSSSPVSGVGVQYKKNDSAPCDNOIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSS 507
DB 327 ---ANTPVSNLKVFEFYNNSPDTTNSINPQFKVTNTGSSAIDLKSLTLRYYYTVDGQKD 383
QY 508 TLVYNCDAAM-----CGGNIRASFGSVNPATPTADTYLQLSFTGTLAAGSGTGE 558
DB 384 QTFW-CDHAAIIGSNGSYNGITSNVKCTFVKMSSTNNADTYLEISFTGTLPEGAHV-Q 441
QY 559 IQRVKNKSDWSNFDNDETNDYSGTNTAFQDWTKVTYVYNGRLVWGTEPSGT 608
DB 442 IQGRFAKNDSNVTQSDNYSFKRSQFVENDQVTAYLNGVLVWGKEPGGS 491

RESULT 13
US-09-670-141-10
Sequence 10, Application US/09670141
Patent No. 6429000
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schuelein, Martin
APPLICANT: Lange, Niels Erik K.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Schnorr, Kirk
TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
TITLE OF INVENTION: Licheniformis
FILE REFERENCE: 5377-200-US
CURRENT APPLICATION NUMBER: US/09/670,141

```

RESULT 14
PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; TITLE OF INVENTION: PROTEINS AND USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```


According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:

Published_Applications_NA contains nucleic acid sequences; the search results will have the extension **.rnpb**.

Published_Applications_AA contains amino acid sequences; the search results will have the extension **.rapb**.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicants submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

Sequences in the PGPub database are public information; it is permissible to leave these results in the case.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:55:11 : Search time 14.7802 seconds
(without alignments)
776.473 Million cell updates/sec

Title: US-09-917-378-1

Perfect score: 4122

Sequence: 1 MGVRPARAFVATAAGTAV.....GFNGSYSGTNTAPTTLCTAS 762

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep:.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep:.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	813	19.7	680	9 US-10-121-032-25	Sequence 25, Appl
2	812	19.7	663	9 US-10-121-032-61	Sequence 61, Appl
3	696.5	16.9	562	10 US-09-981-900B-5	Sequence 5, Appl
4	494	12.0	427	9 US-09-850-982B-2	Sequence 2, Appl
5	415.5	10.1	493	12 US-10-072-152-12	Sequence 12, Appl
6	245	5.9	1367	10 US-09-801-368-108	Sequence 108, Appl
7	238	5.8	2344	10 US-09-815-242-12713	Sequence 12713, A
8	218.5	5.3	5179	10 US-09-922-217-1068	Sequence 1068, Ap
9	218.5	5.3	5179	10 US-09-833-263-1068	Sequence 1068, Ap
10	199.5	4.8	371	10 US-09-739-861A-1	Sequence 1, Appl
11	199.5	4.8	371	10 US-09-795-583-1	Sequence 1, Appl
12	198.5	4.8	386	10 US-09-739-861A-5	Sequence 5, Appl
13	198.5	4.8	386	10 US-09-795-583-5	Sequence 5, Appl
14	189	4.6	596	12 US-10-006-867-100	Sequence 100, App
15	189	4.6	596	12 US-10-052-586-310	Sequence 310, App
16	186	4.5	560	10 US-09-815-242-13057	Sequence 13057, A
17	183.5	4.5	502	10 US-09-815-242-5904	Sequence 5904, Ap
18	182.5	4.4	957	10 US-09-922-217-1065	Sequence 1065, Ap
19	182.5	4.4	957	10 US-09-833-263-1065	Sequence 1065, Ap

175 4.2 1333 10 US-09-815-242-10936 Sequence 10936, A
21 172.5 4.2 688 10 US-09-864-761-36047 Sequence 36047, A
22 164 4.0 113 10 US-09-864-761-39507 Sequence 39507, A
23 163.5 4.0 503 10 US-09-801-368-424 Sequence 424, App
24 163.5 4.0 605 10 US-09-801-368-428 Sequence 428, App
25 162 3.9 881 10 US-09-816-860A-2 Sequence 2, Appl
26 161.5 3.9 802 10 US-09-823-240-2 Sequence 2, Appl
27 159.5 3.9 1031 10 US-09-815-242-10932 Sequence 10932, A
28 154 3.7 1723 10 US-09-841-132-394 Sequence 394, App
29 154 3.7 1723 10 US-09-841-132-395 Sequence 395, App
30 152.5 3.7 945 8 US-08-965-272-2 Sequence 2, Appl
31 152.5 3.7 945 12 US-10-121-882-2 Sequence 2, Appl
32 149.5 3.6 832 10 US-09-729-674-176 Sequence 176, App
33 148.5 3.6 659 12 US-10-090-624-12 Sequence 12, Appl
34 148 3.6 572 10 US-09-943-075A-2 Sequence 2, Appl
35 147 3.6 1140 12 US-10-124-557-104 Sequence 104, App
36 147 3.6 1363 12 US-10-124-557-52 Sequence 52, Appl
37 147 3.6 1404 12 US-10-124-557-2 Sequence 2, Appl
38 147 3.6 1404 12 US-10-124-557-62 Sequence 62, Appl
39 146.5 3.6 2843 8 US-08-681-219-32 Sequence 32, Appl
40 145 3.5 553 10 US-09-888-224-2 Sequence 2, Appl
41 144.5 3.5 1744 9 US-10-108-605-25 Sequence 25, Appl
42 144 3.5 1016 12 US-10-007-693-95 Sequence 95, Appl
43 143.5 3.5 467 10 US-09-863-547B-1 Sequence 1, Appl
44 143.5 3.5 693 10 US-09-752-639-154 Sequence 154, App
45 143.5 3.5 693 10 US-09-984-198-154 Sequence 154, App

ALIGNMENTS

RESULT 1

US-10-121-032-25
: Sequence 25, Application US/10121032
: Patent No. US2002015550A1
: GENERAL INFORMATION:
: APPLICANT: Bylina, Edward J.
: TITLE OF INVENTION: GLYCOSIDASE ENZYMES
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gray Cary Ware & Freidenrich LLP
: STREET: 4365 Executive Drive, Suite 1600
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10121,032
: FILING DATE: 09-Apr-2002
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/09/134,078
: FILING DATE: 13-AUG-1998
: APPLICATION NUMBER: 08/949,026
: FILING DATE: 10-OCT-1997
: APPLICATION NUMBER: 60/056,916
: FILING DATE: 06-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hallie, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 858/677-1456
: TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 680 amino acids
: TYPE: amino acid

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-10-121-032-25

Query Match          19.7%; Score 813; DB 9; Length 680;
Best Local Similarity 41.5%; Pred. No. 2.le-37;
Matches 173; Conservative 65; Mismatches 135; Indels 44; Gaps 13;

Oy 19 AVAAATLGSITMPSATAAPAGFTASGGQFVLNGLPYRYGGTNNYLSYOSHADVDDVL 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12 SVSAEFLLLIVELSFVLFASDEFKVKENGKALNGKEFRFSGNNYMHYKSNMGIDSVL 71

Oy 79 AKQAAMNLSVIRTWGTFIDIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDG----PTG 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 72 ESADMGKIKVIRMGF-----LDGESYCRDKN-----TYMHPEPGVGVPEGISNAQSG 120

Oy 135 LQGLDYAIASAAHGLRIVVLTNDWKEFGMGDOYDKWGLPYHDNFYDPTQOAYKNW 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 FERLDYTVAKAKELGIKLVIVLNNWDFGGMNQYVRWFGGTHDDFYRDEKIKEEYKY 180

Oy 195 VNHLNRYNSITGVYKNDPTIFAWELANEPKRCVSGTLPTSGTCTQATIVNWDOMSA 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 VSFLVNHVNTYGVYREPTIMAWELANEPKRC-----ETDKSGN-----TLVEWVKEMSSY 232

Oy 255 VKSIDPNHMSVSGDEGF---YIGSTQSGS---WPYNDPDSGDVNNALLRVKNIDFCTYHL 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 IKSLDPNHLVAVGDEGFFSNTEGPKYCGEAEWAYNGWS-GVDWKKLLSIETVDFGTFHL 291

Oy 309 YPNYWGQN----ADMGTQIKDHIANAAAIKPTILEEFG--WQTP-DRDSVYQVWTQTV 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 YPSHWGVSPENYAQWGAKEWIEDHIAKEIGKPVVLEEYGIKPSAPVNRRTAIYRLNLDLV 351

Oy 362 RTNCEAGNFWMLAGNVGNOP-----YPNYDGFNVYPSSTATVLAASE-ALAI 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 YDLGGGANFWMLAGIGGSDRDERGYYPDYDGFRIVNDSDPEALIREYAKLFNTG 408

RESULT 2
US-10-121-032-61
; Sequence 61, Application US/10121032
; Patent No. US2002015550A1
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEO for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/121.032
; FILING DATE: 09-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347

```

```

;
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 61:
US-10-121-032-61

Query Match          19.7%; Score 812; DB 9; Length 663;
Best Local Similarity 42.8%; Pred. No. 2.3e-37;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

Oy 41 FVTASGGQFVLNGLPYRYGGTNNYLSYOSHADVDDVLAKQAAMNLSVIRTWGTFIDIGSL 100
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17 FVKVENGKALNGKEFRFSGNNYMHYKSNMGIDSVLESARDMGKIKVIRMGF-----L 71

Oy 101 DGSVPTIDGNKNGFYFOYWDPTGAPAYNDG---PTGLQGLDYAIASAAHGLRIVVVL 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 72 DGESYCRDKN-----TYMHPEPGVGVPEGISNAQSGFERLDYTVAKAKELGIKLVIVL 125

Oy 157 TNDWKEFGMGDOYDKWGLPYHDNFYDPTQOAYKNVNNHLLNRYNSITGVYTKNDPTI 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 VNNWDFGGMNQYVRWFGGTHDDFYRDEKIKEEYKYVSLVNHVNTYGVYREPTI 185

Oy 217 FAWELANEPKRCVSGTLPTSGTCTQATIVNWDOMSAVYKSIDPNHMSVSGDEGF---YI 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 MAWELANEPKRC-----ETDKSGN-----TLVEWVKEMSSITKSLDPNHLVAVGDEGFFSNTE 237

Oy 274 GSTQSGS---WPYNDPDSGDVNNALLRVKNIDFGTYHLYPNYWGQN---ADMGTQIKD 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 GFKPYGGEAEWAYNGWS-GVDWKKLLSIETVDFGTFHLYPSHWGVSPENYAQWGAKEWIED 296

Oy 327 HIANAATAIKPTILEEFG--WQTP-DRDSVYQVWTQVTRTNGEAGNFWMLAGNVGNOP- 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 HIKIAKEIGKPVVLEEYGIKPSAPVNRRTAIYRLNLDLVYDLGGGANFWMLAGIGGSDR 356

Oy 383 -----YPNYDGFNVYPSSTATVLAASE-ALAI 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 DERGYYPDYDGFRIVNDSDPEALIREYAKLFNTG 391

RESULT 3
US-09-981-900B-5
; Sequence 5, Application US/09981900B
; Patent No. US20020138878A1
; GENERAL INFORMATION:
; APPLICANT: Sticklen, Masomeh B
; APPLICANT: Maqbool, Shahina B
; APPLICANT: Dale, Bruce E
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULOSE TO FERMENTABLE SUGARS
; FILE REFERENCE: MSU 4.1-539
; CURRENT APPLICATION NUMBER: US/09/981.900B
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 60/242,408
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-981-900B-5

Query Match          16.9%; Score 696.5; DB 10; Length 562;
Best Local Similarity 27.3%; Pred. No. 3.6e-31;
Matches 221; Conservative 93; Mismatches 194; Indels 301; Gaps 34;

```


[illegible]

RESULT 8
US-09-922-217-1068
; Sequence 1068, Application US/09922217

```

: Patent No. US20020076414A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Lodes, Michael
: APPLICANT: Secretist, Heath
: APPLICANT: Benson, Darin
: APPLICANT: Meagher, Madeleine
: APPLICANT: Stolk, John A
: APPLICANT: Wang, Tongtong
: APPLICANT: Jiang, Yuxiu
: APPLICANT: Smith, Caroline
: APPLICANT: King, Gordon
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan
:
: TITLE OF INVENTION: COMPOSITIONS OF MATTER FOR THE TREATMENT OF CANCER
:
: FILE REFERENCE: 201021.47
:
: CURRENT APPLICATION NUMBER: 20020076414A1
: CURRENT FILING DATE: 20020910
: NUMBER OF SEQ ID NOS: 112
:
: SOFTWARE: FastSeq for Win.
:
: SEQ ID NO 1068
:
: LENGTH: 5179
:
: TYPE: prt
:
: ORGANISM: Homo sapiens
:
: US-09-922-217-1068

```

Query Match	5.3%	Score 218.5;	DB 10;	Length 5179;
Best Local Similarity	22.6%	Pred. No. 0.00029;		
Matches 98;	Conservative	69;	Mismatches 159;	Indels 107;
				Gaps 19;

Qy	394	PSSTATVLA	SEALAI	STGTSPP	--SPSSPSSSP	-----SPSPASPSA	-----S	439
Db	1657	PSPTTTPSP	--PI	TTTTTTP	PTTTPSSPI	TTTTPSPTT	TTTTPSPTT	1714
Qy	440	PSASSSP	SPSS	-----SPVSG	CKVQYKND	SAP--GDNQ	IKPGLQ	488
Db	1715	PSSTTSP	PTTTP	PTTTP	PTTTP	TTTTP	TTTTP	1774
Qy	489	VDLST	-----VTVRYW	-----FTR	DGSGSL	VYN-C--DWA	AGCGN	532
Db	1775	TTTPTTP	CPVPLC	NWGLD	SGKPN	FKPGD	TELIGD	1829
Qy	533	ATPTADY	LQLS	FTGTLA	AGSGT	GEIQNR	VKNKSD	587
Db	1830	DVP	-----IGOLG	TGVCDV	SVGLIC	NEDOK	PGGVI	1882
Qy	588	WTKVTVV	NGRLV	MGTPE	SGTSP	PTPSP	PTPSP	636
Db	1883	TTMTTTT	TENP	TPPTT	PTTTT	TVTP	PTPTG	1942
Qy	637	SPSP	-----SPSP	PTPSP	SPSSP	SVSSG	VCGRAT	689
Db	1943	TTTPTT	TTTTV	TPPT	PTGOT	PTTPT	TTT	1984
Qy	690	SRATSGM	TVAM	SGGNO	TVYN	WNTAL	TQSGA	749
Db	1985	TQPTT	PTT	-----TTT	TVTP	PTPTG	OTPTT	2033
Qy	750	GTNTAP	L	TCTAS	762			
Db	2034	PTTPTT	TTTTT	2046				

```

RESULT 9
US-09-833-263-1068
: Sequence 1068, Application US/09833263
: Patent No. US20020110547A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: APPLICANT: Stolk, John A.
: APPLICANT: Meagher, Madeleine J.

```

```
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-833-263-1066

Query Match      5.3%  Score 218.5; DB 10; Length 5179;
Best Local Similarity 22.6%; Pred. No. 0.00029;
Matches 98; Conservative 69; Mismatches 159; Indels 107; Gaps 19;

Qy 394 PSSTATVLASALAICTGTSPPP--SPSSPSSSSPSP-----SPSPSASPSPA-----S 439
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1657 PSPTTTTTPSP--PITTTTTPPTTTPSSPITTTTTPSPPTTMTTTPSPPTTTPSSPITTTT 1714
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 440 PSASSPSPSPSS-----SPVSGVKVQYKNDSDAP-GDNQIKPGLQLVNTGSSS 488
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1715 PSSTTTSPSPPTTTPSPPTTTPSPPTTMTTLPPTTSSPLTTTLPSPPTTTPSPFS 1774
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 489 VDLST---VTVRYW-----FTRDGSSTLYYN-C--DMAAMCGCNIRASFGSYNP 532
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1775 TTPTTTPCPLCNNTGWLDSCKPFPKPGDTELIGDVGCPGMAA----NISCR-ATMYP 1829
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 533 ATPTADTYLQLSFTGGTLAAGSGTGEIONRVNKS DWSNFDNDYSYGTNTW-----AFQD 587
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1830 DVP-----IGQJGQTVVCDVSVGLICKNEDQKPGGVPMAFCLNVEINVCCECVTP 1882
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 588 WTKVTYVYNGRLVMTGTPSGTSPSPSPSPSP-----SPTPSPSSSPSP 636
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1883 TTMTTTTENTPPTTPITTTTTPPTPTGTPTTPTTITTTTTPPTPTPTGTGTP 1942
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 637 SPSP-----SPSPSPSPSPSPSPSSVSSGVCRAHYVYVNSDMGSGFTATVTVNTG 689
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1943 TTTPTTTTTPPTPTGTPTTPTTITTT-----TVTPTPTG 1984
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 690 SRATSGMTVAWSFGNGTNTVYMWNTALTSQASVTATNLSYNNVLOPCOSTTFGFNGSYS 749
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1985 TOTPTTPTIT-----TTTTVTPTPTGTGTPTTPTTPTTPTTPTPTPTPTPTPTPTPT 2033
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 750 GTNTAPTITCTAS 762
   | | | | |
Db 2034 PTPTPTPTTPTTPT 2046
   | | | | |

RESULT 10
US-09-739-861A-1
; Sequence 1, Application US/09739861A
; Patent No. US2002007692A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Brien E.
; APPLICANT: van der Kleij, Wilhelmus A.H.
; APPLICANT: van Solingen, piet
; APPLICANT: Weyler, Walter
; APPLICANT: Goedegebuur, Frits
; TITLE OF INVENTION: No. US2002007692A1 Cellulase Producing Actinomycetes,
; FILE REFERENCE: GC540-3
; CURRENT APPLICATION NUMBER: US/09/739, 861A
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: US 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: US 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 371

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-833-263-1066

Query Match      4.8%  Score 199.5; DB 10; Length 371;
Best Local Similarity 20.8%; Pred. No. 0.00021;
Matches 101; Conservative 49; Mismatches 153; Indels 183; Gaps 22;

Qy 303 FGTVHLXPNYWGQADWGTOWIKOHIAAAIGKPTILEEFCWQTPDRDSYVQTWTQTVR 362
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39 YGTTTIQRIYVQNNRMGT-----SATQCIN 64
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 363 TNGEAGNFMWLAGNV--NGQP--YPN-YDGFNVYPSSTATVLAASEALAICTGTSPPPS 417
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 VTGN-GFEITQADGVSPTNGAPKSYPSYVDGCH-YGNCAPTTTL-----PM 108
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 418 PSSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 466
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109 RISSIGSAPS-----SVSYRYTGVYNAAYDIWL 138
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 467 DSAPGDNOIKPGLQLVNTGSSSVLDLSTVTVRYWFTRDGSSSTLVYVNCDAAMCGCNIRAS 526
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139 DPTPTNGV-----NRTIIMWFR----- 158
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 527 FGSVNPATPTADTYLQLSFTGTLAAGSGTGEIONRVNKS DWSNFDNDYSYGTNTAF 585
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159 ---VGPVQPIGSPV-----GTAHVGRSWEV-----WTGNSGNDVISFLAPSAI 200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 586 QDWT-KVTYVYNGRLVWG-TEPSGTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 643
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201 SWSFVDKDFVDQAVSHGLATPDWYLTISIAGFEPMEGTGLAVNSFSASVAGCGNCGT 260
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 644 P-TPSPSPSPSPSPSPSSVSSGVCRAHYVYVNSDMGSGFTATVTVNTGSRATSGMTVAWSF 702
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 261 PGTA-----ACQVSYSYTHI-WPCGFTVDITITNTGSTPVDGWELDTL 303
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 703 GGNQTVNWMNTALTSQASVTATNLSYN-NVIOGQSTTTFGNGSYSGTN-TAPT----- 756
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 304 PAGHTVTSVMNALISPASGAVTARSTGSGNRIANGGTQSPFGQCTSSCAGTAPAGARL 363
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 757 --LTCT 760
   | | | | |
Db 364 NGTSCT 369
   | | | | |

RESULT 11
US-09-795-583-1
; Sequence 1, Application US/09795583
; Patent No. US20020081702A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: No. US20020081702A1 Cellulase Producing Actinomycetes,
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/795,583
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 371
```

;
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" - Streptomyces
; OTHER INFORMATION: thermobaceous
US-09-795-583-1

Query Match 4.8%; Score 199.5; DB 10; Length 371;
Best Local Similarity 20.8%; Pred. No. 0.00021;
Matches 101; Conservative 49; Mismatches 153; Indels 183; Gaps 22;

QY 303 FCTYHLYPNWQONADWGTOWIKDHIAAAIGKPTILEEFGWQTPDRDSVYQVTQTVR 362
DB 39 YGTTTIQDRYVQNNRWGT-----SATQCIN 64
QY 363 TNGEAGNFWMLAGNV--NGQP--YPN-YDGFNVYPSSTATVLALEALAIATGTSPPPS 417
DB 65 VTGN-GFEITQADGVSPTNGAPKSPSVYDGCII-YGNCAPRTTL-----PM 108
QY 418 PSSSPSSP 466
DB 109 RISSIGSAPS-----SVSYRYTGNVYNAAYDIWL 138
QY 467 DSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVWFTRDGSSTLVYNCDAAMCGNIRAS 526
DB 139 DPTPTNGV-----NRTEIMWFNR-----158
QY 527 FCSVNPATPTADTYLQLSFTGCTLAAGSGTGEIQNRVKNKSDNSFDETND--YSYGTNTAF 585
DB 159 ---VGPVQIGSPV-----GTAHVGGRSWEV-----WTGSGNSNDV1SFLAPSAI 200
QY 586 QDWT-KVTYVNGRLVMG--TEPSGTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 643
DB 201 SWSFVDFKDFVQVASHGLATPDWYLTISIAGFEPWEGGTGLAVNSFSSAVNAGGNGCT 260
QY 644 P-TPSP 702
DB 261 PCTPA-----ACQVSYSTHT-WPGGFTVDTTITNTGSTPVDGWELDFTL 303
QY 703 GGNQTVNTWNTALTQSGASVTATNLVSYN-NVLPQCSSTFGFNGSYSGTN-TAPT---- 756
DB 304 PAGHTVTSVWNAALISPASCAVTARSTGNSGRIAAANGCTQSGFQGTSSGAGTAPAGARL 363
QY 757 ---LTCT 760
DB 364 NCTSCT 369

RESULT 12
US-09-739-861A-5
; Sequence 5, Application US/09739861A
; Patent No. US20020076792A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Brian E.
; APPLICANT: van der Kleij, Wilhelmus A. H.
; APPLICANT: van Solingen, Piet
; APPLICANT: Weyler, Walter
; APPLICANT: Goedegebuur, Frits
; TITLE OF INVENTION: No. US20020076792A1 Cellulase Producing Actinomycetes,
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
; FILE REFERENCE: GC540-3
; CURRENT APPLICATION NUMBER: US/09/739,861A
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: US 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: US 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 386

;
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" - Streptomyces
; OTHER INFORMATION: thermobaceous
US-09-739-861A-5

Query Match 4.8%; Score 198.5; DB 10; Length 386;
Best Local Similarity 20.8%; Pred. No. 0.00025;
Matches 101; Conservative 49; Mismatches 153; Indels 183; Gaps 22;

QY 303 FCTYHLYPNWQONADWGTOWIKDHIAAAIGKPTILEEFGWQTPDRDSVYQVTQTVR 362
DB 54 YGTTTIQDRYVQNNRWGT-----SATQCIN 79
QY 363 TNGEAGNFWMLAGNV--NGQP--YPN-YDGFNVYPSSTATVLALEALAIATGTSPPPS 417
DB 80 VTGN-GFEITQADGVSPTNGAPKSPSVYDGCII-YGNCAPRTTL-----PM 123
QY 418 PSSSPSSP 466
DB 124 RISSIGSAPS-----SVSYRYTGNVYNAAYDIWL 153
QY 467 DSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVWFTRDGSSTLVYNCDAAMCGNIRAS 526
DB 154 DPTPTNGV-----NRTEIMWFNR-----173
QY 527 FCSVNPATPTADTYLQLSFTGCTLAAGSGTGEIQNRVKNKSDNSFDETND--YSYGTNTAF 585
DB 174 ---VGPVQIGSPV-----GTAHVGGRSWEV-----WTGSGNSNDV1SFLAPSAI 215
QY 586 QDWT-KVTYVNGRLVMG--TEPSGTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 643
DB 216 SWSFVDFKDFVQVASHGLATPDWYLTISIAGFEPWEGGTGLAVNSFSSAVNAGGNGCT 275
QY 644 P-TPSP 702
DB 276 PCTPA-----ACQVSYSTHT-WPGGFTVDTTITNTGSTPVDGWELDFTL 318
QY 703 GGNQTVNTWNTALTQSGASVTATNLVSYN-NVLPQCSSTFGFNGSYSGTN-TAPT---- 756
DB 319 PAGHTVTSVWNAALISPASCAVTARSTGNSGRIAAANGCTQSGFQGTSSGAGTAPAGARL 378
QY 757 ---LTCT 760
DB 379 NCTSCT 384

RESULT 13
US-09-795-583-5
; Sequence 5, Application US/09795583
; Patent No. US20020081702A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: No. US20020081702A1 Cellulase Producing Actinomycetes,
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/795,583
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Unknown

APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052,586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2002, 08:50:51 ; Search time 3966 Seconds
(without alignments)
16796.854 Million cell updates/sec

Title: US-09-917-378-2

Perfect score: 2289

Sequence: 1 atgggtcctagtcgctgcgcc.....tcacctgcacggcctagtlga 2289

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl :
1: gb_ba :
2: gb_htg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_om :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :
29: em_vi :
30: em_htg_hum :
31: em_htg_inv :
32: em_htg_other :
33: em_htg_mus :
34: em_htg_pln :
35: em_htg_rtd :
36: em_htg_mam :
37: em_htg_vrt :
38: em_sv :
39: em_htgo_hum :
40: em_htgo_mus :
41: em_htgo_other :

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	205	9.0	3237	1	AF200304	AF200304 Caldicell
2	180.4	7.9	3004	1	ACU33212	U33212 Acidithiobac
3	180.4	7.9	3004	6	AX467594	AX467594 Sequence
4	180.4	7.9	3004	6	123701	123701 Sequence 6
5	172.8	7.5	4567	1	AF163837	AF163837 Caldicell
6	133.6	5.8	11707	1	AF078737	AF078737 Caldicell
7	133.6	5.8	11707	6	E35099	E35099 Truncated c
8	128.4	5.6	6005	1	AF078042	AF078042 Caldicell
9	128.4	5.6	6416	6	E35100	E35100 Truncated c
10	126.4	5.5	1992	6	AR205120	AR205120 Sequence
11	126.4	5.5	2043	1	TMA17980	Y17980 Thermotoga
12	126.4	5.5	2043	6	AR205088	AR205088 Sequence
13	124.8	5.5	20990	1	AE001779	AE001779 Thermotog
14	124.4	5.4	2001	1	AY033477	AY033477 Thermotog
15	124.4	5.4	2001	1	TNE17981	Y17981 Thermotoga
16	122	5.3	5437	1	CASR69XYN2	AF036924 Caldicell
17	118.6	5.2	1265	8	ASNMANNA	L35487 Aspergillus
18	118.6	5.2	1302	6	AR023984	AR023984 Sequence
19	116.4	5.1	4977	1	CDCMANA	L01257 Caldocellum
20	114.8	5.0	2977	6	A28170	A28170 B. laetus st
21	113.2	4.9	2831	1	BACCELA	M76588 Bacillus la
22	110.6	4.8	7686	1	AF038547	AF038547 Bacillus
23	106.2	4.6	5439	1	CDCCELA	L32742 Caldocellum
24	102.8	4.5	4743	1	AB008029	AB008029 Thermophi
25	100.8	4.4	5513	1	AT286105	Z86105 A. thermophi
26	98.4	4.3	3262	1	AF078038S1	AF078038 Caldicell
27	98.2	4.3	5284	1	CSU16708	U16708 Caldocellum
28	97.8	4.3	1621	1	THFE2AA	M73321 Thermomonos
29	97.8	4.3	3509	1	PAE488923	AJ488933 Paenibac
30	92.2	4.0	1434	8	ABCELA4MR	Z50095 Aquarius bi
31	90.2	3.9	2029	6	E35142	E35142 Truncated c
32	90.2	3.9	2029	6	E35143	E35143 Truncated c
33	88.4	3.9	210319	2	AL607086	AL607086 Mus muscu
34	87.8	3.8	40104	1	SCF11	AL132662 Streptomy
35	87	3.8	165165	2	AC096270	AC096270 Rattus no
36	86.4	3.8	37941	1	AOPKZA361	AJ223998 Amycolato
37	86.4	3.8	41906	1	SC5C7	AL031515 Streptomy
38	85.8	3.7	4161	1	BSP133614	AJ133614 Bacillus
39	85.4	3.7	3480	1	CF1END	L02544 Cellulomora
40	85.2	3.7	2801	5	AF218784	AF218784 Gallus ga
41	85.2	3.7	4241	1	CSECB	X13602 Caldocellum
42	84.6	3.7	163636	10	AL669937	AL669937 Mouse DNA
43	84	3.7	207070	10	AC122843	AC122843 Mus muscu
44	83.8	3.7	185241	2	AC115962	AC115962 Mus muscu
45	83.6	3.7	108360	14	H510LR	D10875 Herpes simp

ALIGNMENTS

RESULT 1
AF200304
LOCUS

DEFINITION

AF200304 3237 bp DNA linear BCT 01-DEC-2000
Caldibacillus cellulovorans putative type IIb cellulose-binding
domain gene, partial cds; and beta-1,4-xylanase xynA precursor
(xynA) gene, complete cds.

ACCESSION AF200304

VERSION AF200304.1

KEYWORDS GI:7385018

SOURCE Caldibacillus cellulovorans.

ORGANISM Caldibacillus cellulovorans

Bacteria: Firmicutes; Bacillales; Alicyclobacillaceae;

Caldibacillus.

REFERENCE 1 (bases 1 to 3237)

AUTHORS Sunna,A., Gibbs,M.D. and Bergquist,P.L.

TITLE	A novel thermostable multidomain 1,4-beta-xylanase from 'Caldibacillus cellulovorans' and effect of its xylan-binding domain on enzyme activity
JOURNAL	Microbiology 146 (Pt 11), 2947-2955 (2000)
MEDLINE	20519260
PUBMED	11065373
REFERENCE	2 (bases 1 to 3237)
AUTHORS	Sunna,A., Gibbs,M.D. and Bergquist,P.L.
TITLE	Direct Submission
JOURNAL	Submitted (29-OCT-1999) Biological Sciences, Macquarie University, North Ryde, Sydney, New South Wales 2109, Australia
FEATURES	location/Qualifiers
source	1..3237
	/organism='Caldibacillus cellulovorans'
CDS	/db_xref='taxon:74586' <1..279 /note='orf1' /codon_start=1 /evidence='not experimental' /transl_table=11 /product='putative type IIb cellulose-binding domain' /protein_id='AA61648.1' /db_xref='GI:7385019' /translation='OVGCNLSKRSVKSLSGRTGADYYIPIFTSCAGSLAPCASGG IOVRINKNDWTNYNEANDSYDPTKTFAIDNNRVTLIRNQQLIWGEV'
gene	384..3149
CDS	/gene='xynA' 384..3149 /gene='xynA' /note='multidomain protein' /codon_start=1 /transl_table=11 /product='beta-1,4-xylanase xynA precursor' /protein_id='AA61649.1' /db_xref='GI:7385020'
	/translation='MGTGRAGEWIKMLLVLAIGLIPYPHVAENVLIIQSDFE DDTTCOGVGGVTELTVTSAAYSGYGLSVSRTEWHGPTLDITSYIQGKTGYLP SAMWLKPGSSSTRI SMTWORTMODTVYEQIVFDTALSNWITQLKAQKLYEPAVNL QEYFPAATHQSIFYIDDVRIEOLPDLTKVEENIPSLKDVFAGREP IGTAFENFELL DOQRKLILKHNSVTPGNVLKWDSTEPOEGVENFTESDKAVAFVQNGMK IRGHTLLI WINOTPNWIFYOSGNLVSKEVLYORMERHIKT VSVRYGKIITYAWDVNEVIDPGQP GLRSLWYOIAGDEYTEKAFQAHEADPNALLFINDTNTHESCOKSALYNLVORKLSK GIPVGHGHTHINIWPSSIETENSLEKFSNI GVVOEITEIDMSITNASSQKYDYL P SDLAOOATYRHOLFEMFLKRSSLIONVTFWCKDDANTWLRFPPVVRNDWPLLI FDELQ KAKPAYAVGVTVSPPTSTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPSASQTL RCVEYRWGSSATDNOMKPOLRI VNTGSQAVPLTELKVRYWYTKNSTQAEQPCDWAQI GYSNIRAOFPVSLAQPGVGDYIELSFTGCSVPACGTCIEIONRIIFTWNMYNETDD WSYNCOTWGPSPSTRILLRNGLVMGTFPGGSSPTPTPTPTPTPTPTPTPTPTPTPT STPT TTTSVPLSELTIRYWTVDGHPQPVFCMDMAWVCNSLRGSLVKLTGRTGADYYILEI TTTSGAGSLAPGANSDDIOARINKNKNDWTNYNEANDSYDPTKTFAIDNNRVTLIRNQ LWGVGEV'
sig_peptide	384..482
mat_peptide	/gene='xynA' 483..3146 /gene='xynA'
misc_feature	/product='beta-1,4-xylanase XynA' 483..965 /gene='xynA' /note:'Region: thermostabilizing domain' /region: 1946..1946 /gene='xynA' /note:'Region: beta-1,4-xylanase catalytic domain' /region: 1947..2075 /gene='xynA' /note:'Region: proline-threonine linker' /region: 2076..2519 /gene='xynA' /note:'Region: cellulose-binding domain typeIIrb' /region: 2520..2696 /gene='xynA' /note:'Region: proline-threonine linker' /region: 2697..3146

REFERENCE	FEATURES	source
Actinomycetales: Frankineae; Acidothermaceae; Acidothermus.		
1 (bases 1 to 3004)		
AYMON, K.A., HIMMEL, M.E. and THOMAS, S.R.		
Direct Submission		
Submitted (04-AUG-1995) Steven R. Thomas, Applied Biological Sciences, National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401, USA		
	Location/Qualifiers	
	1..3004	
	/organism="Acidothermus cellulolyticus"	
	/strain="11B (ATCC 43068)"	
	/db_xref="taxon:28049"	
stem_loop	710..725	
RBS	762..769	
repeat_region	781..817	
	/note="three direct repeats of ACGGGGCA"	
	/rpt_type=direct	
	/rpt_unit="781..788	
CDS	824..2512	
	/codon_start=1	
	/transl_table=11	
	/product="E i beta-1,4-endo glucanase precursor"	
	/protein_id="AA75477.1"	
	/db_xref="GI:988300"	
	/translation="MPRALRRVPGSRVMLRVGVVAVLVAALANLAVPRPARAAGG GYWITSORELIDANNYPVRITAGINWECFETCNVVICLWSRDYRSMLDQIKSLGYNTI RLPYSDILLKPTMPNSINIFYONNODLGITSLOVMDKIVAYAGQIGIRIILDDRHRPD CSQSAALNRTSSVSEAFNIDSLQALAQRYKGNPTVVGFDI.HNEHPDAPACWCGPDSID WRLAARAGHNAVLSVAPNLLIIFVEGVSYNGDSYWMGNI.QGACQYPPVILNDRPRLV SAHDYATSVYPTQTFSDPTFPNNMPTGIIWNKNGYLFNQNTAPVMLGEFGTTLOSTLDQ TWLTKLVQYLLKPTAFQACDSFOWTFWSNPDSGDTGTCLIKDDMQTVDTYKDGXLYLPIK SSIFDPGASGASPCSPSPSPSPSASRTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT ASPT TNSWNAAVTONGQSVTARNMSYNNVTPQCGNTTTFGRGTYGSAALPTVCAAS"	
sig_peptide	847..946	
mat_peptide	947..2509	
	/product="E i beta-1,4-endo glucanase"	
	/EC_number="3.2.1.4"	
	/function="cleaves beta-1,4-glycosidic linkages of cellulose"	
	/note="cellulase"	
stem_loop	2514..2560	
BASE COUNT	604 a 938 y 531 t	
ORIGIN		
Query Match	7.9%	Score 180.4; DB 1; Length 3004;
Best Local Similarity	64.8%	Pred. No. 3.5e-15;
Matches 284; Conservative	0; Mismatches 151; Indels	3; Gaps 1;

[illegible]


```
/protein_id="AAD30363.1"
/db_xref="GI:4836167"
/translation="MKRRVSIILSLFLFLINTLVGTLIFHQEAKAAATYVDFEGADTL
SYQPAFGASIVADMGATNGKSSVRVSNRSSIWDGAVDVKNIMNGTTWVVSAYVKH
IIVSEVVDYNGDIQIDMDSNLSNAFTSSGFEQSTEGQWARGSVTKVQSLTWORRFA
GKSLYVSGRTSNHIGAIQIPVDITLQEGGVYKISVWYVNSGSKQKMSLTWRRFATD
PSTSVENLILNROVPSNTWMLSGSYSLPAGVTYSELLLYVEAONANLAPWDDLLKY
DLSKLAEPEWEIPLSLEKRYDYFKGVGVALSYKLSASDTEKKKWLKHFNSITAGNEMKP
SELLVDENTYNFESKADERVFNATSNIAIRGHTLVNHEQTPDMFFKGTNGTSLKDAL
LSRLKQYITLVYNGRTKGVYAVDVEAIDESQDGFRRSNMYNCSPEYIEKAFIWA
HEADPDLFYVNDYNTENSQRQFIYNNIKSLKEKGVPIHGTLQSHINLQDSYSEI
ENTIRLFSSIPLGLEIHIITELDMSEYQMSYSPTRDILLIKOAMRYKELDLFKKY
NNVITSYTFMLKDDYSWLSQNFSGDPLLDENYKSKYAFWSLLEPTVIPANSTLP
APPALQIPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
PTNVLNPKIKIENVGTTAVDLRSVRKYRYWYTTIDGQTSQSVTSISIPAYIDVAPVKL
GANAGGADYVEIYFGKSGAGVLAAGOSTKEIRLSIQKSGSYNOSNDYSVRSATGYIE
NEKVTYIDVVLVWGREPSRNAQIKVYIANGNLSPTNLNPKIKIENVGTTAVDLRS
VRYWYTTIDGQTSQSVTSISIPAYIDVRFVKLGANAGGADYVEIYFGKSGAGVLA
AQOSTKEIRLSIQKSGSYNOSNDYSVRSATGYIENEKVTGYIDCAIYVNGREPSRGT
PAGVVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
PTPTPTPTGTCGSLKLVLYKNNETSASTSSIRPWPKIYVNGSGSSVDLSRVKIRYKY
TVDDGKPOSACVDNAQIGASNVTFNFYKLSGVSGADYILEVGFSSGAGQLOPKDGT
DIQVRFNKNDSNNTQADDWSLOSMTNGENAKVTLYVDCVLVWQEGQLOPKDGT
TATPTPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
FCADPAVLVYCGRYMYLTNDILEYDENGVKDNYSYKINKITVYISDDLVMWTDHGF
IEVAGPNGLKAWSLAPAVACKKINGKDRFELYFGNSGGIGVITADSTPGWMSDP
LGRPLITWSTPCVOCVWMLFDPVAVLDDGKRAYLYFGGVGPQGDAMPNTARVWGLD
MISVWGSVAVTIPAPMEFEDSGINKIGNTYYSYCTNFAQRPQSPAGAIAYWTGRS
PMGWHEYKCVILIRNPNFGVGGNNHQLFEPNCKWYIAYHAQTLKDLGVAKGYRSP
HINIVQIENGITIKVYADYKGVAKNDFPDMVEAETFAWCAGIISTKAKANNNMCL
TGISDQWIALSLVDFGNAGPKQFEAOVSSINGKGYIELRIDSDVDGRTIAVAEVLPOS
GSSQWQKVEANVENVTGVIIDLYLIFRGEKKSNLFDMDWREVR"
6255..10535
/gene="celb"
6255..10535
/gene="celb"
/feature="note"
/note="multidomain enzyme with an amino-terminal glycosyl
hydrolase family 10 domain, and a carboxy-terminal
glycosyl hydrolase family 5 domain"
/codon_start=1
/transl_table=11
/product="CelB"
/protein_id="AAD30364.1"
/db_xref="GI:4836168"
/translation="MKRRVLRVSRLLIIVFIMSLVSGMSYFPVKTEAAPDWSIPS
LWESYKNDKFGVAPARCLSDNDKQWLVKFNSTIAENEMKPESELLAGOTSTGLSY
RFSTADTFVNFVANNINIGIRGHTLVWNIQTPDMFPROSSGOMLSKDALLRKQYIYD
VGRYKGVYANDVNNVNEAIDESQPDGYRSTWYQICGPEYIEKAFIWAHEADPNAKLF
YNDYNTLSIKRDFIYNNKNIKSKGVPIHIGIMOSHINWNPVSVEISENISKLFSSI
PGIEIHTLELDMSLYNYGSNENYSTPTODLLIQRAQYKDIFTMLRYKGIIVTCVTEW
GLKDDYSWLNSSSKRDMPLLFDDYSAPKPAYWSYIEAAGASAPSPVTATPTPTPTPT
TVTVATPTPTPTGTCGSLKLVLYKNNETSASTSSIRPWPKIYVNGSGSSVDLSRVK
IRYWTYVDDGKPOSACVDNAQIGASNVTFNFYKLSGVSGADYILEVGFSSGAGQLOP
KGDAGDLOVRNKNKDNWSYNQADDWSLOSMTDYGENAKVTLYVDCVLVWQEGQLOP
PAPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
WTPSESGLAKVWYANGNLSPTNLNPKIKIENVGTTAVDLRSVRKYRYWYTTIDGQTS
QSVASVSNIPAYIDVILGANAGGADYVEIYFGKSGAGVLAAGOSTKEIRLSIQKSG
SYNOSNDYSVRSATGYIENEKVTGYIDVILVKG..SVTSSINPAYIDVFKVFLCANAG
NPKIKIENVGTTAVDLRSVRKYRYWYTTIDGQTSQSVTSISIPAYIDVFKVFLCANAG
GADYVVEIYFGKSGAGVLAAGOSTKEIRLSIQKSGSYNOSNDYSVRSANSYIENEKVT
GYIDGAIWNRREPSKCYKPAQVVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
KPMW..VYSHDQGLVYQQPWFYPGTGTLYNDQWRNWTYIMDNGIAPLLI
nPLAPALUK..SKRGFVGLDHRKRLPLGTNGKNINITYYQNGEKKPPVPRKN"
11145..111707
/feature="Orf2"
/codon_start=1
```

```
/transl_table=11
/product="unknown"
/protein_id="AAD30366.1"
/db_xref="GI:4836170"
/translation="MSFLKQYQNSLKMILQIISDALKKMTDSMURKDYFEGYKNT
NVFVPPCAFESSNFPEKALDSYMDKSRKIKTILILDTNPTYNKAIDIVDSVEKVLKN
SLEFVDTKFTFEGVGGISSNHDLSIYFKDRTLRKLIWISILLIMPTISNSIFNAA
AVVAIVFDIYVALSILITEMIFKIPN"
BASE COUNT 3828 a 1994 c 2994 g 2891 t
ORIGIN
Query Match 5.8%; Score 133.6; DB 1; Length 11707;
Best Local Similarity 52.4%; Pred. No. 5.3e-09;
Matches 395; Conservative 0; Mismatches 344; Indels 15; Gaps 4;
QY 1220 GAATCAGTACCGGCGCATCGCCTCCGCTCGCGAGCTCGAGTCCATCCTCTCTCTCT 1279
DB 3883 CAACACACCTCACCAGCATCAGCCCCACACAGCCAGCCACACGTCAGCAGCAACGCTCA 3942
QY 1280 CTCCTGTCGGCGCTCTCCGTCGGCGTCTCCCTCGCGCTCTCCGTCGGCGCTCA 1349
DB 3943 CTCCTAACCGCGGAGCGAGTACGCTTACTGTGACTCCGACACCGACACACACACA 4002
QY 1340 GCCCGCTCTCCGTCGTCGTCGCCGCTGTCGCG---GTGGCGTGAAGTGCATACAGAACA 1396
DB 4003 CGCCGACACCGACAGGACACCTGCGACCGGAAGTCGTTTGAAGGTACTATACAGAACA 4062
QY 1397 ATGATTCGGCGCGGCTGATACACAGATCAACCGCGGCTCTCAGTTGGTGAATACCGGCT 1456
DB 4063 ATGAGACAAGTCGCGACACAAGTCTTATAAGCCGCTGGTTTAAAGATAGTCAATGCAAGCA 4122
QY 1457 CGTCGTCGCTGGATTTGTCGACGGTGCACGCTCGGCTACTGCTTACCCCGGCTGCTGGCT 1516
DB 4123 GCAGCAGTGTTCATCTTACGAGGTTTAAGATAAGTACTGTTACACAGTGGATGGTG--- 4179
QY 1517 CTTGCACACTGTGTACAACTGTGACTGGCGCGCATGGGCTGTGGGAATATCCGCGCT 1576
DB 4180 ACAAGCCACAGAGTGGCTGATGCTGACTGGCCACAGATAGGGGCAAGCAATGTGCATTTCA 4239
QY 1577 CTTTCGGCTCGGTGAACCGCGGACGCGCGGACACCTACCTGCTGCTGCTGCTGCTTCA 1646
DB 4240 ATTTGTTGAAGCTGAGCAGCGGAGTGAGTGGAGCGGATTTATTTACTTTGGAGGTAGGATTTA 4299
QY 1637 CTGCTGGAAAC-----GTTTGGCGCTTGGTGGGTGCAACCGCTGAGATTTCAAAACGGCTCA 1690
DB 4300 GCAGTGGAGCTGGGCAAGTTGCAGCCTGCTTGAAGACACAGGGGATATACAGGTAAAGTTTA 4459
QY 1691 ATAAGAGTGAATGGTGGAACTTTTGTATGAGACCAATGACTACTGCTATGAGGAGCAACACCG 1750
DB 4360 ACAAGAATGACTCGAGCAATTAACAATCAGCCAGACGACTGGTCTGCTGCTGCTGCTGCTGCT 4419
QY 1751 CCTTCCAGGATTTGGA--CGAAGGTGACGGTGTATGCTCAATGGCGGCTGGTCTGGGCTCA 1807
DB 4420 CGAATATGAGAGAAATGGGAAGGTAAACCTGTATGCTAGATGGTGTCTTCTGGTATGGGGG 4479
QY 1808 CTGAACCTCTGGGCAACCGCCACCGCCACACCGCCAGCCCAAGCCCAAGCCCGCTCGCCCA 1867
DB 4480 AGGAGCGGGCGGAGGACACCTGCACCGACAGCAAGCAGACACCAACCGCCCACTCCCA 4539
QY 1868 GCCCAGCGCCGACCCCAAGCCCGAGCTCTCTCCCATCCCGCTCCCGGAGCCCGCAGCCCA 1927
DB 4540 CAGCAACAGCAACACCGCGCGGACAGCAACCGCAACGCTCTACACCGACCGCAGCAAC 4599
QY 1928 GCCCTACCGCGCTCCCGCTCGCGGAGCCCGCTCGCC 1961
DB 4600 CACCAACCCCAATACCAACACCCCAACCGCTCC 4633
RESULT 7
LOCUS E35099
DEFINITION Truncated cellulase composition.
ACCESSION E35099
PAT 18-JUN-2001
```

E35099.1 GI:13018924
JP 1999221086-A/1.
SOURCE
ORGANISM
REFERENCE:
AUTHORS
TITLE
JOURNAL
COMMENT

Palji.A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,
Hyu,M. and Baian,P.W.
Truncated cellulase composition
Patent: JP 199221086-A 1 17-AUG-1999;
CLARIANT INTERNATIONAL LTD
OS Unknown
PN JP 1999221086-A/1
PD 17-AUG-1999
PF 21-SEP-1998 JP 1998283606
PR 19-SEP-1997 US 08/932571
PI PALJI ANDERSON,PETAER L BAGUKUISUTO,ROY M DANIEL, PI
GURAHAMU K PAKINTON,
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONIONOTISU WILLIAM
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42//C12N1/21,C12R1:19), PC
(C12N9/42,C12R1:19),C12N15/00

CC Key Location/Qualifiers
FT source 1..11707 /organism='Unknown'.
FT Location/Qualifiers
1..11707
/organism='unidentified'
/db_xref='taxon:32644'

HASE COUNT 3828 a 1994 c 2994 g 2891 t
ORIGIN

Query Match 5.8%; Score 133.6; DB 6; Length 11707;
Best Local Similarity 52.4%; Pred. No. 5.3e-09;
Matches 395; Conservative 0; Mismatches 344; Indels 15; Gaps 4;

QY 1220 CAATCAGTACCGGCACATCGCTCCCGCTCCCGCAGCTCGAGTCGATCCATCCTCGTCCGCGT 1279
DB 3883 CAACACACACCTGCACCGACATCAGCCCCGACACCGAGCCACAGTGCAGCAGAACGCCGA 3942
QY 1280 CTCCTCGCTCGCTCGCTCGGCGTCTCCGCTCGCGCTCTCCGTCGCGCTCTCGTCGCCGA 1339
DB 3943 CTCCAACCGCGACCGACAGTACGCGTTACTGTACTCCGACCGACACCAACACCCGA 4002
QY 1340 GCCCTCTCCGCTCGCTCGCTCGCGCGTGTGCGG---GTGGGCTGAAGTGCAGTACAAGACA 1396
DB 4003 CGCCGACACCGACAGGACACCTGGCAGCGGAAGTGGTTGAAGGTACTATACAAGACA 4062
QY 1397 ATGATTCCGCGCGCGGTGATACACAGATCAACCGCGTCTCCAGTTGGTGAATACGGGT 1456
DB 4063 ATGACAGAAGTGCAGACACAGTTCTATAAGCCCGTGGTTTAAGATAGTGAATGGAGCA 4122
QY 1457 CUTCCTCGCTGGATTGTCTGACGCTGACGCTGCGGTGCTGCTTACCCGGGATGGTGGT 1516
DB 4123 CGACGAGTTGATCTTACGACGGTTAAGATAAGATACTGTTACAGATGGATGGTGGT--- 4179
QY 1517 GCTGCACACTGGTGTACAACTGTGACTGGCGCGCATGGGTGTGGGAATATCCCGCGCT 1576
DB 4180 ACAAGCCACAGTACGCGGTATGTGACTGGGACAGATAGGCGCAAGCAATGTGACATTCA 4239
QY 1577 CTTTCGCGCTCGGTGAACCGCGCGACCGCGACACCTACCTGCAGTTGCTGTCA 1636
DB 4240 ATTTTGTGAAGTGCAGACGGAGTGAAGTGGAGCGGATTTACTTTGAGGTAGGATTGA 4299
QY 1637 CTGTTGGGAAC-----GTTGGCCGCTGGTGGGTGCGAGCGGTGAGATTCAAAACCGGGTGA 1690
DB 4300 CGACTGGAGCTGGGCAGTTGCGACCTGTGAAGACACAGGGATATACAGGTAAGGTTTA 4359
QY 1691 ATAAGAGTACTGGTGCAGACTTTGATGAGACCAATGACTACTCTGATGGGAGCAACCG 1750
DB 4360 ACAAGAATGACTGGAGCAATTAACAATCAGGACGACGATGGTGTGTCGAGACATGA 4419
QY 1751 CTTCCAGGATTGGA---CGAAGGTGACGGTCTATGTCAATGCCCGCGCTGGTGTGGCGGA 1807

Db 4420 CGAATTTATCGAGAGATCGAAGCTAAGCTGTATAGATGCTGTCTGCTATGGGCGC 4479
QY 1808 CTGAACCGTCCGGCAGCAGCCCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1867
Db 4480 AGGAGCCGGCGGAGCGACACCTTGCACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4539
QY 1868 GCCCGAGCCCGACCCCAAGCCAGCTCTCCCATCCCGTCCCGGAGCCCGAGCCAGCCAGC 1927
Db 4540 CAGCAACAGCAGCAGCAGCCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4599
QY 1928 GCCCTACGCGTCCCGTCCCGCAGCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCC 1961
Db 4600 CACCAACCCCAATACCAACACCCACACACGCTCC 4633

RESULT 8
AF078038S5
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AF078038S5 6005 bp DNA linear HCT 11-FPH-2001
Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial
cgs; and Cele gene, complete cds.
AF078042
AF078042.1 GI:12743878
5 of 5
Caldicellulosiruptor sp. Tok7B.1.
Caldicellulosiruptor sp. Tok7B.1
Bacteria; Firmicutes; Clostridia; Clostridiales;
Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE 1 (bases 1 to 6005)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the
extreme thermophile Caldicellulosiruptor isolate Tok7B.1
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)
MEDLINE 201711169
PUBMED 10706665
REFERENCE 2 (bases 1 to 6005)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
Sydney, NSW 2109, Australia

FEATURES
source
1..6005
/organism='Caldicellulosiruptor sp. Tok7B.1'
/isolate='Tok7B.1'
/db_xref='taxon:80339'
c1..506
/note='ORF6; multidomain'
/codon_start=3
/transl_table=11
/product='glycosyl hydrolase 6'
/protein_id='AAK06393.1'
/db_xref='GI:12743884'
/translation='GQPDITNMNLTGPNLHVKIYDVGTDLGITASLANALLYSAAT
KEYGVSDEAAKNLAKELLDWMNLTGPNLHVKIYDVGTDLGITASLANALLYSAAT
PBGDIKSGVKFIDIRSKYKODPDQKLVSAYNAGEAEFRHFRWAQCDIAIANTY
EILFNGQ'
misc_feature
c1..503
/note='Region: family 48 glycosyl hydrolase domain'
617..5872
/note='multidomain glycosyl hydrolase'
/codon_start=1
/evidence=experimental
/transl_table=11
/product='Cele'
/protein_id='AAK06394.1'
/db_xref='GI:12743885'
/translation='MOENKAIKRVVVSITALLVTLSLCFPGIMPVKAYAGGTYNNGEA
LQRTIMFYFQNSKRLPSVNRNMRGSDGDDGLDGLTGGWHADGHDHFNLPMS
YSASMLGWAVYEKDAFYKSKOLEHILNQIEWANDYFYKCHPSKYVYVYQVGDPTVDH
NFWGPAEYVHOMKRPAYKCDLSNPASSVVAETAASLAVASVVIKERNOKAASYLQHAK

Db 2593 ACACCAACAGTACACGAACGCCGACCCGACCTCCCTACACCGACACCGGCGGTACACCT 2652
QY 1363 GTCTCGG---GTGGGTGAAGTGCAGTACAGAAACAATGATTTGGCGCGCGGUTGATAAC 1419
Db 2653 GGTACGGGAAGTGTGTGAAGGTACTGTACAAACAACAATGAGACAAGTCCGAGCAGAGT 2712
QY 1420 CAGATCAAAACCGGCTCTCCAGTTGGTGAATACACGGGCTGCTGCTGGTGGATTGTCGACG 1479
Db 2713 TCTATAAGCGGTGGTTTAAGATAGTGAATGGAGGAGCAGCAGTGTGTGATCTTAGCAGG 2772
QY 1480 GTGACGCTCGGTACTCGTTTACCCGGGATGGTGGGTGCTGCGACACTGGTGTACAACCTGT 1539
Db 2773 GTTAAGATAAGTACTGGTACACAGTGGATGCTG---ACAAGCCACAGAGTGGGTATGT 2829
QY 1540 GACTGGCGCGGATGGGTGTGGGAATATCCGCGCCTCGTTCCGGCTGGGTGAACCCGGCG 1599
Db 2830 GACTGGGCACAGATAGGGGCAAGCAATGTGACATTTCAATTTTGTGAAGCTTAGCAGCGGA 2889
QY 1600 ACCCCGACGGCGCACACCTACCTGCAAGTTGCTGCTACCTGGTGGAAAC---GTTGGCC 1653
Db 2890 GTGAGTGGACGGATTTATACCTGGAGGTAGGATTTAGCAGTGGAGCTGGGCGAGTTCCAG 2949
QY 1654 GCTGGTGGTTCGACGGGTGAGATTCAAAACCGGGTGAATAAGAGTGAAGTGGTGGAACTTT 1713
Db 2950 CTTGGTAAGCACACAGGGGATATACAGTAAGCTTTAAACAAGATGACTGGACCAATTAC 3009
QY 1714 GATGACACCAANTGACTACTGCTATGCGACGACACACCGCCTTCCAGGATTGGA---CGAAG 1770
Db 3010 AATCAGGACAGACCACTGGTGTGATGCGAGAGCATACCAATTTATGAGAGAATGCGAAG 3069
QY 1771 GTGACGCTGTATGTAATGCGCGCTGGTGGGGACATCAACCGTCCGCGACACGCCCC 1830
Db 3070 GTGACGCTGTATGTAAGTGTCTGCTGATGGGCGCAGGACCGGAGGAGCGACACCT 3129
QY 1831 AGCCGACACCGCAGCCGACGCAACCGCTCCCGAGCGCGGACCGGACCCCAAGCCCC 1890
Db 3130 GCACCGACAGCAGCAGCAACACCAAGCGCAACTCCGACAGCAACCCCAACACCTACACCT 3189
QY 1891 AGCTCTCCCTCCCTCCCGTCCCGGACCGCCAGCCGCTACGCGCTCCCGCTGCGCG 1950
Db 3190 ACACCGACCCGACACCGAGTGTGTAAGCCCAACACCGCGCAGCGGATCACCG 3249

RESULT 10
LOCUS AR205120 1992 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 57 from patent US 6368844.
ACCESSION AR205120
VERSION AR205120.1 GI:21502622
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE UnClassified.
AUTHORS 1 (bases 1 to 1992)
BYLINE E.J.
TITLE Glycosidase enzymes
JOURNAL Patent: US 6368844-A 57 09-APR-2002;
FEATURES Location/Qualifiers
source 1..1992
/organism="unknown"
BASE COUNT 600 a 405 c 573 g 414 t
ORIGIN

Query Match 5.5%; Score 126.4; DB 6; Length 1992;
Best Local Similarity 49.3%; Pred. No. 6.5e-08;
Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;

QY 121 TTCTGTCACCGCATCCGCGGCTCAGTTGCTGTAACGGCTTCCCTATCTACGGGCGGA 180
Db 49 TTCTGTAAGTGGAAACGGAATTCGCTCTGAACGGGAAGAATTCAGATTCATTGGA 108
QY 181 AGCAACAACCTATTACCTCAGCTATCAGTGCACGCGGACGCTGATGACGCTGTGGCCAAG 240

Db 109 AGCAACAACCTACTACATGCACTACAGAGCAACCGGAATGATAGACAGTGTCTGAGAGT 168
QY 241 GCTCAACGCGATGAATCTTTCTGTCTATCCGACCTCGGGTTCATCTCGACATCGGCTCTCTT 300
Db 169 GCCAGAGACATGGGTATAAAGGTCTCTCAGAACTCTGGGTTCTCTCGACGGGACAGTTAC 228
QY 301 GACGGTCCGTGCGCCACAAATCGAATGCAACAAGAACGGCTTCTACTATTTTCACTACTGGA 360
Db 229 TGCAGAGACAAAGAACACCTACATGCATCTCTGAGCCGCTGTTTTTCGGGGTCCCAAGAAGA 288
QY 361 CCGTCCGACCGCGCTCCGGCTACAAACGACGCGCCGCTTGCAGGCGCTTGCAGCTAC 420
Db 289 ATATCGAACGCC-----CAGAGCGGTTTCGAAAGACTCGACTAC 327
QY 421 CGCATCCGACGCGCGCGCGCGCTTTCGGGTGATTTGCTGCTCTCAACCAACGACTTGG 480
Db 328 ACAGTTCCGAAACGGAAGAAGTCCGTTATAAAGTCTGCTCATTTGTTCTGTAACAACTGG 387
QY 481 AAGAATTTGGGGGAATGGATCAATACGACAAGTGGTACGGCTTCTTACCAGGACAAC 540
Db 388 GAGGACTTCCGTGGAATGAACCACTAGCTAGGTGGTTCGAGGAAACCCATCACGAGCAT 447
QY 541 TTCTACACCCACCCCGCGACCGAGCGGTACAAGAATTTGGTCAATCATCTACTCAAC 600
Db 448 TTCTACAGAGTACAGAAGATCAAGAAGATCAAAAGTACGTTCTCTTCTCTGTAAC 507
QY 601 CGGTTCAACAGCATTTACCGCGGTGACGTACAAGAAGTCAACGATCTTTGCTTGGGAA 660
Db 508 CATGTCAATACCTACACGGGAGTTCTTACAGGGAAGAGCCCACTATCGCTCGGAG 567
QY 661 CTTGCCAATGACGCGCGCTCGGTAGGAGCGGCACATTTACCAACCTCGGCGACCTGTGAC 720
Db 568 CTTTCAACCAACCGCGCTGTGAGACGGAACAAATTCGGGGAACACGCTGTTGAGTGGGTG 627
QY 721 CAGCGGACCATTTGTCAACTGGGTGATCAAAATGTCCGCGTACGTCAAAAGCATAGACCT 780
Db 628 AAGGAGA-----TGACCTCTACATAAAGTCTGATCCCAACCACTCGTGGCTGTG 681
QY 781 AACCATATGCTCTGCTGGCGACGAAGGTTTCTACATTTGGGTCAACGCGAGGAAGCTGC 840
Db 682 GGGGACGAAGGATTTCTACGCAACTACGAAGGATTCAAACCTTACGGTGGAGAGCCGAG 741
QY 841 TGCGCATACAAACACCGCTCCGAGCGGTCCGACAACAATGCTTTCTCCGTGTCAGAAC 900
Db 742 TGGGCTTACAAACCGCTCGTGC---CGGTGTTGACTGGAAGAAGTCTCTTCCATAGAGACG 798
QY 901 ATTGACTTTGGCAGCTATACCTGTACCCGAATTTACTGGGG-----CCACAAC 948
Db 799 GTGACTTGGCAGCTTCCACCTCTATCCGCTCCCACTGGGGTGTCTAGTCCAGAGAATAT 858
QY 949 GCGGACTGGGGAACGAATGATCAAGGATCATATTGCGGAATGCCGAGCGGATCGGCAAG 1008
Db 859 GCCCAGTGGGGAGCGAAGTGGATAGAAGACCAACATTAAGATCGCAAAAGAGATCGGAAGA 918
QY 1009 CCGACCATTTCTCGAAGAATTCGG-----CTGGCAGACACGGGACCGGATTCGGTCT 1059
Db 919 CCGGTTGTTCTGGAAGAATATGGAATTTCCAAGAGTCCGCGCAGTTAACAGACGGCCATC 978
QY 1060 TATCAGACGTGGACCCAGACTGTGCGTACGAACGCGTCAACGAGCTGGAATCTTCTGATG 1119
Db 979 TACAGACTTGGAAACGATCTGGTCTAGGATCTCGGTGGAGATGGAGCGGATGTTCTGATG 1038
QY 1120 CTGCTGGGAATGTCAACGG 1139
Db 1039 CTCGCGGAATCGGGAAG 1058

RESULT 11
LOCUS TM17980
DEFINITION Thermoloba maritima manB gene.
ACCESSION Y17980
VERSION Y17980.1 GI:6006594

KEYWORDS beta-mannosidase; manB gene.
SOURCE Thermotoga maritima.
ORGANISM Bacteria; Thermotoga maritima
REFERENCE 1 (bases 1 to 2043)
AUTHORS Parker,K.N., Lam,D., Duffaud,G., Sincad,M.A., Mathur,E.J. and Kelly,K.M.
TITLE Amino Acid Sequence of beta-mannosidase genes from the hyperthermophilic bacteria Thermotoga maritima and Thermotoga neapolitana
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2043)
AUTHORS Parker,K.N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) K.N. Parker, North Carolina State university, Department of Chemical Eng. Box 7905 NCSU, Raleigh, NC 27695-7905, USA
FEATURES Location/Qualifiers
 1..2043
 /organism="Thermotoga maritima"
 /strain="MSB8"
 /db_xref="taxon:2336"
 1..2043
 /gene="manB"
 1..2043
 /codon_start=1
 /transl_table=11
 /product="beta-mannosidase"
 /protein_id="CAB56854.1"
 /db_xref="GI:6006595"
 /translation="MGIGDDSDSPSVAEFLILLIVELSPVLFASDFVKVNGKFFAL
 NKEFRFGISNNYIMHYKNSGMDIVLESARDMGIKVLRIWGLDGSYCRDKNTYMH
 PEPGVFGEGISNAQSGFEKLDITRIPDQHEMFLEGHFOGKTVKDSIKAKVYNEAR
 GTHHDDVDFSRPEKVEKYYVSLVNVHTVTPGYREEPTIMAWELANEPGCEYDKS
 GNTLVNWKWESSYIKSLDPNHLVAVGDEGFFSNYEGFPGYGGAEWYNGWGVDMK
 KLLSIEVDFCTFHLVPSHWGVSYPENYAOWCAKWIEDIKIAKEIGRPVLEUYGIPK
 SAPVNRATIRLMDLVLDLGGDMFWMLAGIGGSDRDERGYYPDYDGRFTVNDSDS
 PEALIRYAKLFTGEDIREDTCSFILPKDGMELIKRTVEVRAGVFDPYSNPTFKLSVK
 VEDLVFENETBILGYGIYGFOLDITRIPDQHEMFLEGHFOGKTVKDSIKAKVYNEAR
 VYLAEEVDFSRPEKVEKYYVSLVNVHTVTPGYREEPTIMAWELANEPGCEYDKS
 VRVAKFERLSECEILKVDIYIPNVEGLKRLRPYVNLPGWVKIGIDMNNANVESAE
 IITPCGKEYRHFHWRJIEFDRTAGCKEHLIGVVGDLHLYDGFIDNVRLRYRTGGM"
 BASE COUNT 609 a 417 c 592 g 425 t
 ORIGIN
 Query Match 5.5%; Score 126.4; DB 1; Length 2043;
 Best Local Similarity 49.3%; Pred. No. 6.4e-08;
 Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;
 QY 121 TTCTGTCACCGCATCGCGGGTCAGTTCTTTGAAACGGCCCTTCCCTCATGTTACGGGGGA 180
 Db 100 TTCTGGAAGTGGAAACGGAAATTCGCTCTGAACCGGAAAGAAATTCAGATTTCATTGGA 159
 QY 181 AGGAACAATATTACCTCAGCTATCAGTCCGACGCGGACGCTCGATGACGCTTTGCCCAAG 240
 Db 160 AGCAACAATCTACTACATGCATCAAGAGCAACGGAATGATAGACAGCTCTCTCGAGAGT 219
 QY 241 GCTCAAGCGCATCAATCTTTCTCTCATCCGACCTCGGGTCTTCATCCACATCGGCTCTCTT 300
 Db 220 GCCAGACATGGGTATAAAGGTCCTCAGATCTCGGGTTCCTCGACGGGGAGACTTAC 279
 QY 301 CAGCGCTCGTCCGACACANTCGATGGCAACAGACGGCTTCTACTTTTCAGTACTGGGAC 360
 Db 280 TCCAGAGACAAGAACACCTACATGCATCTCTGAGCCCGGTGTTTCGGGGTCCCAAGGA 339
 QY 361 CCG..CCAGCGCGGCTCCGGCTTACAACGACGGCGGACGCTTCCAGGCTTGCAGCTAC 420
 Db 340 ATATCGAAGGCC-----CAGACGGGTTCGAAAGACTTCAGCTAC 378
 QY 421 GCGATCCGCGACGGCGGCGCGACGGCTTCGGGTGATTTGTCGCTCTCCACCAACGACTCG 480
 Db 400 TATATCGAAGGCC-----CAGACGGGTTCGAAAGACTTCAGCTAC 378

Db 379 ACAGTTGCCAAACGCGAAAGAACTCGGTATAAAACCTTTGTCATTCTTCTGTGAACAACCTGG 438
 QY 481 AAAGAAATTTGGGGGAATGATCAANTACGACAACTGCTAGCGGCTTCTCTTACCACGACGAAC 540
 Db 439 GACGACTTCGGTGGGAATGAACCACTAGCTGAGGTGGTTTGGAGGAACCCATCAGACGAT 498
 QY 541 TTCTACACGACGACCCCGGACCGAGCGGTACAAAGGCTTACAAATTTGGGTCAATCATCTACTCAAC 600
 Db 499 TTCTACAGAGATGAGAAGATCAAGAAGAGTACAAAAGTACGCTCTCTTCTCGTAAAC 558
 QY 601 CGGTCAACAGCATTTACCGCGGTGACGTACAGAAGAGTCCAAAGATCTTTGGTTGGGAA 660
 Db 559 CATGTCAATACCTACACGGGAGTTCTTACAGGGAAGAGCCCACTCATGGCTCGGAG 618
 QY 661 CTTTGCAATGAGCGGCTCGGTAGGAGCGGCACATTACCAACCTTCGGGACGTCGCACT 720
 Db 619 CTTGCAACCAACCGCGCTGTGAGACGGACAATTCGGGGAACACGCTCGTTGAGTGGT 678
 QY 721 CAGGCGACCAATTTCTCAACTGGGTTCGATCAAAATTCGCGGTAGCTCAAAACATAGACCTT 780
 Db 679 AAGGAGA-----TGAGCTCTTACATAAAGAGTCTCGATCCCAACCCCTCGTGGCTGT 732
 QY 781 AACCATATGCTCTCGGTGGCGACGAAGGTTCTACATTTGGGTCAACGGAAGGAAGGAG 840
 Db 733 GCGGACGAAGGATTTCTTACGAACCTACCAAGGAATTCAAACCTTACCGTGGACAAGCCGAG 792
 QY 841 TGGCCATACAACGACCGCTCCGACGCGCTCGACACAAATGCTCTTCTCTCGGTGTCGAAGAC 900
 Db 793 TGGGCTTACAAACGCGTGGTCT---CGGTGTGACTGGAAGAAGCTCTCTTTCGATAGAGAG 849
 QY 901 ATTGACTTTGGCAGCTATCACCTGTACCCGAATTTACTGGGG-----CCAGAAC 948
 Db 850 GTGACTTTCGGCAGCTTCCACCTCTATCGTCCCCTGAGTCCAGAGAAGTAT 909
 QY 949 GCGGACTGGGGAACGCAATTTGGATCAAGGATCATATTGCCAATGCCCGCAGCGGATCGGCAAG 1008
 Db 910 GCCCAGTGGGGAGCAAGTGGATGAAGACCAACATTAAGATCGCAAAAGAGATCGGAAA 969
 QY 1009 CGGACCATTTCCGAAGAATTCGG-----CTGGACACACCGGACCGGATTCGGTC 1059
 Db 970 CCCGTGTTCTGGAAGAATATGGAATTCGAAGAGTCCGCCAGTTAACAGACGGGCATC 1029
 QY 1060 TATCAGACTGGACCCAGACTCTGGCTACGAACGCTGGAAGCAGGCTTGGAACTTCTCGATG 1119
 Db 1030 TACAGACTCTGGAAGCATCTGGTCTACGATCTCGGTGGAGATGGAGCGATGTTCTCGATC 1089
 QY 1120 CTCGCTGGGAATGTCACCG 1139
 Db 1090 CTCGCGGGAATCGGGAGG 1109

RESULT 12
 LOCUS AR205088
 DEFINITION Sequence 11 from patent US 6368844.
 ACCESSION AR205088
 VERSION AR205088.1 GI:21502582
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2043)
 AUTHORS Bylina,E.J.
 TITLE Glycosidase enzymes
 JOURNAL Patent: US 6368844-A 11 09-APR-2002;
 FEATURES location/Qualifiers
 1..2043
 /organism="unknown"
 BASE COUNT 609 a 417 c 592 g 425 t
 ORIGIN

Query Match 5.5%; Score 126.4; DB 6; Length 2043;
 Best Local Similarity 49.3%; Pred. No. 6.4e-08;

Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;

QY	121	TTCTGTCACCCATCCGGCGGTGAGTTTCGTCTGAACGGCTTCCCTCATGTTACGGGGA	180
Db	100	TTCTGCGAAAGTGGAAACGGAATTCGCTCTCAACGGGAAGAATTCAGATTCATTGGA	159
QY	181	ACGAACAACATATACCTGACCTATCAGTCCGACGGCGAGTCGATGACGTGTGGCGCAAG	240
Db	160	ACGAACAACATATACCTGACCTATCAGTCCGACGGCGAGTCGATGACGTGTGGCGCAAG	219
QY	241	GCTCAAGCCATGAATCTTCTGTCATCCGACCTGGGTTTTCATCCACATCGGCTCTCTT	300
Db	220	GCAGAGACATGGGTATAAAGTCTCTCAANTCTGGGTTTCTCCAGCGGAGAGATTAC	279
QY	301	GACGGTCCGTCGCCCAATCATGATGCGCAACAAGACGGCTTCTACTTTTCAGTACTGGGAC	360
Db	280	TCGACAGACAAGAACACCTACATGCTCTGAGCCCGGTGTTTCGGGGTCCGAGAAGGA	339
QY	361	CCGTCGACCGCGCTCCGGCGGTACACAGCAGCGGCCGCGGTGCAAGCCCTTGACTAC	420
Db	340	ATATCGAAGCC-----CAGAGCGGTTTCGAAAGACTCGGACTC	378
QY	421	GGGATCCGACGGCGCGCCGCGGCTTCGGGTGATTGTCCTCTCACCACGACTCG	480
Db	379	ACAGTTGCCAAGGAAGAACTCGGTATAAACTTGTCTATGTTCTGTGAACAACCTGG	438
QY	481	AAAGAAATTTGGGGATGATCAATACGACAAGTGTGACGGCTTCCTTTACCACGACAAC	540
Db	439	GACGACTTCGGTGAATGAACAGTACGTGAGGTGTTGGAGGAACCCATCAGCAGCAT	498
QY	541	TTCTACACCGACCCCGGACCGAGCGGTACAGAAATTTGGTCAATCATCTACTGAAC	600
Db	499	TTCTACAGAGATGAGAAGATCAAGAAGATACAAAAGTACGTCTCTTCGTCTAAAC	558
QY	601	CGGTCACACGACATTTACCGCGCTGAGCTACAAGAAGATCCAAACGATCTTTGCTGGGAA	660
Db	559	CATGTCATACCTACACGGGAGTTCCTTACAGGAAGACCCACCATCATCGCTGGGAG	618
QY	661	CTTGCAATGACCGCGGTGCTAGAACGGCGACATTTACCACCTCGGCGACGTGCAC	720
Db	619	CTTGCAAGCAACCGCGCTGTGAGACGGACAATCGGGAACACCGCTCTTCAGTGGGTG	678
QY	721	CAGGCGACCATTTGCAACTGGCTCGATCAAAATGTCGGCTACGTCAAAACGATACACCT	780
Db	679	AGGAGA-----TGAGTCTTACATAAAGACTCTGGATCCCAACACCTCTGCTGTG	732
QY	781	AACCATATGCTCGGTGGCGACGAAGGTTTACATTTGGGTCAACGCGGGAAGCGGC	840
Db	733	GGGACGAAGGATTTCTCAGCAACTACGAAGGATTCAAACCTTACGTCGGAACCGGAG	792
QY	841	TGGCCATACAAGCACCGGTCCGACGGGTGCAACAATGCTCTTCTCGGTGTCAAGAAC	900
Db	793	TGGGCTACACGCGGTGCT---CGGTGTTGACTGGAAGAAGCTCTTTCGATAGAGCG	849
QY	901	ATTGACTTTGGCACCTATCACCTGTACCGGAATTAAGTCTGGG-----CCAGAAC	948
Db	850	GTGACTTCGGGACGTTCCACCTCTATCCGTCGCCACTGGGTGTGAGTCCAGAGAACTAT	909
QY	949	CGGACTGGGAAACCAATGGATCAAGATCATATTGGCAATGCCGCGACGATCGGCAAG	1008
Uc	910	CGCCAGTGGGACGAAGTGGATACAAGACCAACATAAAGATCGCAAAAGAGATCGGAAA	969
QY	1003	CCGACCATTCGGAAGATTCGG-----CTGGGACACCGGACCGGATTCGCTC	1059
Db	970	CCGCTGTGTTCTGGAAGATATGGAAATCCAAAGAGTGGCGCCAGTTTAAAGAACGCGCATC	1029
QY	1060	TATCAGACCTGACCCACACTCTCGCTACGAACGCTGAAGCAGGCTGGAACCTCTCTGATG	1119
Db	1030	TACAGACTCTGGAAGATCTGGTCTACCATCTCGGTGGAGATGGAGCGATGTTCTGGATG	1089
QY	1120	CTCGCTGGGAATGTCAACGG 1139	
Db	1090	CTCGCGGAATCGGGGAAG 1109	

RESULT 13
AE001779/c
LOCUS
DEFINITION
Thermotoga maritima section 91 of 136 of the complete genome.
ACCESSION
AE001779
VERSION
AE001779.1
GI:4981777
KEYWORDS
SOURCE
ORGANISM
Thermotoga maritima.

Thermotoga maritima.
Bacteria; Thermotogae; Thermotogales; Thermotogaceae;
Thermotoga.
1 (bases 1 to 20990)
Nelson, K.E., Clayton, R.A., Gilli, S.R., Gwinn, M.L., Dodson, R.J.,
Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,
Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
Eisen, J.A., Fraser, C.M. et al.
Evidence for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima
Nature 399 (6734), 323-329 (1999)
99287316
-10360571

2 (bases 1 to 20990)
Nelson, K.E., Clayton, R.A., Gilli, S.R., Gwinn, M.L., Dodson, R.J.,
Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,
Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..20990
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
complement(97..1203)
/gene="TM1224"
complement(97..1203)
/gene="TM1224"
/note="similar to PID:1208894 SP:Q44406 percent identity:
54.86; identified by sequence similarity; putative."
/codon_start=1
/transl_table=1
/product="transcriptional regulator, XylR-related"
/protein_id="AAD36299.1"
/db_xref="GI:4981778"
/translation="MPKSVRAENISRIILKRIKSPVSRVELAEELGLTKTTVEIAKI
FLEKGLVVEEDSPKGVGRTPKSLKISPCAYVLGIEVTRDEIAACLDASNNLAHE
AHPLSQSDREELNVMRTIDRAKMMKLGKUSALTVAAPGPIIDTGRGIIIDPRN
PPLSOPLANLLKKEYGIEVVDADMGKVEKWTTRKDSFAWILTKGIGAGII
DGELYRGENGAGEIGYTRFNGNEYVLEDCNENVLKHLVLSMGFSLSAARDSDG
VVRVEYFDIARIYFISGLNLHLPGISKIVIGGFFKELGENFLKKIKIEVTHLLYK
HSDMSFSKVOEPIVAFCAVAHALENLYERTVS"
complement(1217..2197)
/gene="TM1225"
complement(1217..2197)
/gene="TM1225"
/note="similar to GB:Pyro_h percent identity: 52.69;
identified by sequence similarity; putative."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAD36300.1"
/db_xref="GI:4981779"
/translation="MKVTEKIPNIPEWERPEGYTPVMRYSKNPICGRNPVKCARV
FNSAVPVNGEFGVFRIDHKNTBPLFLHFGSKDGIWHEIEPEEIQHVDVNGEPPQPS
YAYDPRVVKIEDTYITFTCDTHGPTICGWTIKDFKTEFLPRLNVPVNRGVLFPK
INGKYVLMNRDNGHTPFQDIFLSESDMTWGNHREVLGRSSYNWENKIKAGPY
PIETSEGMLLIYHGVTLTCNGYVYFSGAALLDLDPSKVLVRSRYLLTPEEVEVTG

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
FEATURES
source

gene
CDS

gene
CDS

FVPNVFPCCALCDADTGRVAIYYGAADTHVALAFGYIDEIVDFVKRNSM"
 complement(2250. .3938)
 /gene="TM1226"
 /complement(2250. .3938)
 /note="TM1226"
 /note="similar to GB:Pyro_h percent identity: 57.09;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="oligopeptide ABC transporter, periplasmic
 oligopeptide-binding protein, putative"
 /protein_id="AAD36301.1"
 /db_xref="GI:4981780"
 /translation="MNMKRLLLVPPVVVLFSSLSFQVLERNETWYCGSLWSPSSNNN
 PFTPDNAVRFTGLVYTFMFFYDPLTGNFDPWLAEGWLDSTKYRVLRREGIYWHDN
 VPLTSDVRFTEIAKRYKGIHYSSVMELDHIETPDNRTVIFVKDPRYHNNELLY
 TLPIVPHKIHIEEKDETTILOSSNEYPLGSGPYVAHSDONKMIPEFENMMGTVMGV
 KPAPYVIVIVRLSNVALGMLKGLDFSNFLKVPYILKKVYNLATWDEPPYHLS
 STVGLFLNARKYPLSLPEPRRAIAMSIADPIVQRYVEGAVLADPLGLPLNSVMK
 YYPKEVVERIGFRYDPEAKSLIDKIGFRDVGNGDGFRETDPGKPIKLTIECPYMTD
 MQAIOIVDOLKVVIGINAEYFPDSSKYIENMYKGEFDIENANGTGISSTPYTFNT
 IFYDALESFSTGYNTGRYONPEVESLLELNRTPLDNVEKVTGLGKGLCEILLKDL
 PFIPLYGAMAFITQDNVNTWMPNEHPYAWPCGANNWQTGALKILFLNLPKAK"
 complement(3935. .5944)
 /gene="TM1227"
 /complement(3935. .5944)
 /note="similar to PID:1679597 percent identity: 54.44;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="endo-1,4-beta-mannosidase"
 /protein_id="AAD36302.1"
 /db_xref="GI:4981781"
 /translation="MRRFMEILLIVELSPVLFASDEFVKVENGKFAALNGKREFRIGSN
 NYMHYKSNRMTDSVIESARDMGIKVLRIWGFIDGESYCRDKNTYMHPEPCVGVPEG
 ISNAQCFERLDYTVAKAKELGKLVILVNNDDFCGMQYVWFGGTHRDQYRDE
 KIKKEYKYVSFLVNHVNTYGVYREEPTIMAWELANEPCECTDKSNTLVYHVKEM
 SSYLKSLDPIHILVAGDEGFSYNGEYKPGGAEAWAYNGSWGVDKRLLSIETVDFR
 TFIPLPSHWGSPENYQAWGAKFIEDHIKAIKEIGKVPVLEEYGIKPSAPVNRATYR
 LWNLDVLDLGGCFWMLACIGESDRDERGYYPDYDGRIVNDDSPAEALIREYAK
 LNTGEDIREDTCSFLLPKDGMETIKTVEYRAGVFDYNTFEKLSRVKVEDLVFENET
 HLGVIYGFGLDTRIPDGEHEMFLEGIHQGTQKDSIKAKVNEARYVLAEEVDFSS
 PEVKNVNSCTQWAEFCSPDIENWGEVGNALQNLVLPCKSDMEARVVARKEFRLS
 ECETLEVDIYIPIVNEGKRLRPLVAVLPGVNLGLDMNANVESAILITFGKREYRR
 FHVIREDRTAGYKELHIGVVGDLHRLDGPFIIDNRLYRTGGM"
 complement(6103. .6903)
 /gene="TM1228"
 /complement(6103. .6903)
 /note="similar to GB:M81878 SP:P26833 PID:144859 percent
 identity: 55.56; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator, Rpir family"
 /protein_id="AAD36303.1"
 /db_xref="GI:4981782"
 /translation="MKIRDKILNVYTFQSPAERKRVADYVLERPDDVLIHYSITTEFAKIV
 GSVETTHRMKILDPFGYQAFKIALARELSGLEETIERRDFIDEEDIDILRLRLDLD
 MKNVEKAVENILSAHRIPLFCVGLSGVSEYASLKLFLGFTFFHDEEDNDPHVQVIEAVN
 LTGDELVISISHTGNI RDYTKQVAKDMGAKTIAITNRQSELAKEVHVLVLOSPPVK
 YDTVEFLRENGEIAVDVLFKETFQRIYRERKKHFPENLEGVFKPRF"
 complement(6887. .8335)
 /gene="TM1229"
 /complement(6887. .8335)
 /note="similar to GB:Pyro_h percent identity: 69.66;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAD36304.1"

/db_xref="GI:4981783"
 /translation="MGSASGSKSQRKAGFAPRVSQIIRYKTSRPELQSKMSMLQK
 ERPGGGRDEFKRRGNFSSPLKSVLSQDCRQTOLRSGKEALQKFKETRIITEGCGIT
 MKVVVGIPSYNNAETISIVATAAGIIVDFDGDGMVNSDGSADGTRERFMEITDF
 GLPESFVEGLPKGSMRAIMEFALKODAEAVVFLADLRSVKPWWVERLAGPVLK
 GEADYVTPYLRHRFDGTITNNVCFTMTAVYLGKVRQPIGGDGVGKHLLEIYIGKP
 KEIWTNDVAREGIDIMWTITTAINESGRVVOAALGTVKHDGPKHILKGMPLVVGTL
 FELVITYENVKKEIWKIHDVPIYGTPOEEVPSMSIDIGNLKLARETLERVEYIDRG
 ILSEVKESGTLSSSWVDTLVRSAYVYRTRDKKVVENLLPYFARTARFAEEVKSLS
 DEAEERYVEOLDVFELEKHSIREEMKVEDNR"
 complement(8026. .9309)
 /gene="TM1230"
 /complement(8026. .9309)
 /note="similar to percent identity: 0.00; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAD36305.1"
 /db_xref="GI:4981784"
 /translation="MKIALIHVRCGLMDGVSLMEKWKVLTMKCHEVHIIVAEKNKEG
 VDLTLKEIGFENPDFERVNRNPFQGIKDFLSEKFEFLDFLKEEELFHLINLAKDYD
 LIVPNNISWGLFSPSLGALSRLKKNFVAHHIDFWMRKHLIPENNRPREIIDKHFPF
 DLPNKHVINTIAORELKRNRNIDSVVPMDFSPITSEMYHRYRELOIAPGT
 IVALQATRIDRRKTIELSDVSLKLTETLSKKEADLYNGERYSEVILFLSGICEDE
 EYLKELKEYASSGVSLLVSEEVKNTSLFKLYNAADFVTPSILKGCWNLLEAI
 AAKKPVLYFEYEFKSDIKPAGLKVYSLGDRCFRNGLVKVDRIILKKAVEEISRLLP
 DPSLYRETVEHNFVYGRHSLESLERLEDILSREVLP"
 complement(9306. .11804)
 /gene="TM1231"
 /complement(9306. .11804)
 /note="similar to SP:P21139 PID:205299 percent identity:
 49.50; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="alpha-mannosidase-related protein"
 /protein_id="AAD36306.1"
 /db_xref="GI:4981785"
 /translation="MKVKKVVNIHNDREWFTSSVTSKMLKEVFRVKELVOKNPEFV
 YVLDQTAAYVEDLLVYHPDLEEDIRELVRSHRIIFGIEKVLWNGISFPNDGISOFFW
 GEKNTGFRMRNAGLLDSFGHISQEPOLHRIIFGIEKVLWNGISFPNDGISOFFW
 KGSDGTAVQGVFLVGGYRNLYNLKETQDIAEKRLKHEVKEIAKFSRGEIILLDGDI
 DLSPEKPDYLNVEIVSEEPERFENAPTLSELGSLGSRVACVPGTFLSTRVYLKLE
 SDLVKLLKMNLLSVLNGEELSLFWRDYLTLAHNSCGVACDAPVHEKMEYRNV
 VFLVKSSVESRLHLVERLGEKGYALSLSPFSYNHHFCDCGERTFLKLTNGAGLKT
 EVADEKKEVKELTWENYRAYFEGDGTLLNLRNKAAGVLKFKEDGDATVTSVETDF
 SISLEKRTVLOTKHSWVLEERTIRSSQCLIEKRVIFDETLPLVWRILAKI KGRN
 YKLSMVTSGSEIFAVKMPDVVKRKKDLYLLPETPEELKGIILJAARETGEVREFF
 FQGEVSYMSESSKTIILARGLREYMEEGKVHVTLRSRVEWIAEKVRNVDGAGPMY
 VPAKCEGVFTDLAFMELPVHPRSEEFRRWYLLFDNPPLLDVDEADCTEKEKTFPFS
 SLPMAGYKDRYHLWVYNYPLLEIEGLRPLQIGEKSIESLSEKQCFVSRNLNLFPEFP
 VVSVEKNPEGEVTEFLKELTSSLEDEITNLGKDDPRSKHRLFSLLRTKKEIELSLQY
 LIDVEKTARENLRLMKRTYDYVVELFESEERE"
 complement(11801. .12868)
 /gene="TM1232"
 complement(11801. .12868)

Query Match	Score	DB 1:	Length	20990:
Best Local Similarity	49.28;	Pred. No. 7.4e-08;		
Matches	512;	Conservative	0;	Mismatches 477; Indels 51; Gaps 5;
Qy	121	TTCTGTCACCGCATCCGCGGTTCAGTTCTGCTGAACGCCCTTCCCTATGCTTACGGGGA	180	
Db	5878	TTCTGTAAGTGGAAACGGAAATTCCTCTGAACGGAAAGAAATTCAGATTTCATTCGA	5819	
Qy	181	ACGAACAACATATTACCTCAGCTATCAGTCGCACGCCGCCGACGTCGATGAGCTGTGTGCCCAAG	240	
Db	5818	AGCAACAACATCTACATGCTACACAGCAAGCAAGATGATAGACAGTGTCTTCGAGAGT	5759	
Qy	241	GCTCAACGGATGAATCTTCTCTCATCGGACCTGGGGTTCATCGACATCGGCTCTCTT	300	

Db	5758	GCAGAGACATGGGTATAAAGGTCTCAGAACTCGGGTTTCTCGACGGGAGAGATTAC	5699
Qy	301	GACGGCTCGTGGCCCAACTCGATGCGCAACAACAGCGCTTCTACATTTCAGTACTCGGAC	360
Db	5698	TCCAGAGACAAGAACACCTACATCGATCTCTGACCGCGGTGTTTCGGGTGCCAAGAGGA	5639
Qy	361	CGGTCCAGCGCGCTCCGGCGTACAAACGAGCGGCCACCGGCTTGAAGCCCTTGACTAC	420
Db	5638	ATATCGAACGCC-----CAGACGGTTTTCGAAAGACTTCGACTAC	5600
Qy	421	CGCATCGGACGGCGCGCGCGGCGCTTCGGGTCATTGTCGTCTCTCACCACGACTGG	480
Db	5599	ACAGTTGCGAAACCGGAAAGAACTCGGTATAAAACATTGTCATTGTTCTTGGAACAACCTGG	5540
Qy	481	AAAGAAATTTGGGGGAATGATCAATACGACAAGTGTACGGGCTTCTCTTACCACGACAAC	540
Db	5539	GACGACTTCGGTTCGAATGAACAGTACGTGAGGTGTTGGAGCAACCCATCACGAGAT	5480
Qy	541	TTCTACACGACCGCCCGGACCGAGCGGTACAGAATTTGGGTCAATCATCTACTGAAC	600
Db	5479	TTCTACAGAGATGAGAAGATCAAAAGAGAGTACAAAAGTACGTCTCTTCTCGTAAAC	5420
Qy	601	CGGTCAACAGCATTTACCGCGGTGAGCTACAAGACGATCCCAAGCATCTTTGCTTGGGAA	660
Db	5419	CATGTCATACCTACACGGGAGTTCTTTACAGGGAAGAGCCCAACCATCATGCGCTGGGAG	5360
Qy	661	CTTGCAATGAGCGCGCTCCGTAGGAAGCGGCACATTACCAACCTCGGCGACGTGCAC	720
Db	5359	CTTGCAAAACGACCGCGCTGAGACGGACAAATCGGGAACACAGCTCGTTGAGTGGGTG	5300
Qy	721	CAGGCGACCATTTCTCACTGGGTGCGATCAAAATGTCGGCGTACGTCAAAAGCATAGAC	780
Db	5299	AAGGAGA-----TGAGCTCTACATAAAGAGTCTGCATCCCAACCACTCTGCTGGCTGTG	5246
Qy	781	AACCATATGCTCGGTCCGCGACGAAGGTTCTACATTGGGTCAACGAGGAAACCGCG	840
Db	5245	GGGGAGGAAGGATTTCTCAGCAACTACGAAGGATTCAAACCTTACGGTGGAGAAGCCGAG	5186
Qy	841	TGGCCATACAAGACCGCTCCGACCGGCTCGACAACATGCTCTTCTCGGTGTCGAAGAAC	900
Db	5185	TGGGCTACAACCGCTGGTC---CGGTGTTGACTCGGAAGAGCTCTTTCGATAGAGACG	5129
Qy	901	ATTGACTTTGGCAGCTATCACCTGTACCCGCAATTACTGGGG-----CCAGAAC	948
Db	5128	GTGGACTTCGGCAGCTTCACACCTTATCGCTCCCACTGGGCGTGTCAAGTCCAGAGAAT	5069
Qy	949	CGCGACTTGGGAACCGCAATCAAGATCATATTGCGAATGCCGCGCGGATCGGCAAG	1008
Db	5068	GCCAGTGGGAGCGAAGTGGATAGAGACCATATAAGATCGAAGAGAGATCGGAAA	5009
Qy	1009	CGACCATTTCTCGAAGATTCGG-----CTGGCAGACACCGGACCGGATTCGGTC	1059
Db	5008	CCCGTTGTTCTGGAAGATATGGAATTCCAAAGAGTGCAGGATTAACAGAACGGCCATC	4949
Qy	1060	TATCAGAGCTGCACCACTGTGCTAGCAACGGTGAAGAGGCTTGGAACTTCTGGATG	1119
Db	4948	TACAGACTCTGGAAGCATCTGCTCTACGATCTCGGTGGAGATGGAGCGATGTTCTGGATG	4889
Qy	1120	CTCGCTGGGAATGTCAACGG	1139
Db	4888	CTCGCGGGAATCGGGAAGG	4869
RESULT	14		
LOCUS	AY033477	2001 bp	DNA linear
DEFINITION	Thermotoga neapolitana		
ACCESSION	AY033477		beta-mannanase (man5) gene, complete cds.
VERSION	AY033477.1		GI:18642981
KEYWORDS			Thermotoga neapolitana.
SOURCE			Thermotoga neapolitana
ORGANISM			Bacteria; Thermotogae; Thermotogales; Thermotogaceae;

REFERENCE	1 (bases 1 to 2001)	Thermotoga.
AUTHORS	Parker,K.N., Chhabra,S.R., Lam,D., Callen,W., Duffaud,G.D., Sneed,M.A., Short,J.M., Mathur,E.J. and Kelly,R.M.	
TITLE	Galactomannanases Man2 and Man5 from Thermotoga species: growth physiology on galactomannans, gene sequence analysis, and biochemical properties of recombinant enzymes	
JOURNAL	Biotechnol. Bioeng. 75 (3), 322-333 (2001)	
MEDLINE	21473892	
PUBMED	11590605	
REFERENCE	2 (bases 1 to 2001)	Thermotoga.
AUTHORS	Parker,K.N., Chhabra,S.R., Lam,D., Callen,W., Duffaud,G.D., Sneed,M.A., Short,J.M., Mathur,E.J. and Kelly,R.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (26-APR-2001) Chemical Engineering, North Carolina State University, Room 3309, 840 Main Campus Drive, Raleigh, NC 27606, USA	
FEATURES	Location/Qualifiers	
source	1..2001	
	/organism="Thermotoga neapolitana"	
	/strain="DSM 5068"	
gene	/db_xref="taxon:2337"	
	1..2001	
CDS	/gene="man5"	
	1..2001	
	/gene="man5"	
	/codon_start=1	
	/transl_table=1	
	/product="beta-mannanase"	
	/protein_id="AAK53459.1"	
	/db_xref="GI:18642982"	
	/translation="MKKLYFSLVTLPIVLFANSDFVKVENCGRFLNCEEFKRVGSN NTYMHKSNRMIDSVLSAKAMGVKRLRWGLDGSYCRDKNTYMHAPVFGFLPEG TAAQGFERLDYVAKAKELGLIILVNWDDDFGCMQYVRFGGIHHDDFYRNEK IKKEYKYVSFLINRVNTYGVYREPTIMAEELANPRCETDKSGNTLVEWVEKMS AYIKSLDPNHLVAVGDGFNNYEGRPYGAEMAYNGMSVDKRLLEIEIVDFCT PHLYPSHMGVSPENVAOMCAKIEDHIKIAKEVGKPVLEEYGIKPSAPNRVAIYKL WDLVYNLGGCAFMFMLAGICEGHDREKGYYPDGFRIYNDESEAKLIREYAKL FSTGDTREDTCHFITPKDGEIKTKTVKRVGVDFYSNTFKGISGVGNLLIFDEIKH LGQVYGFEDTTRISDGEHMFLEAHFGETVKDITRVKVNRAQYVLAHEVDFSRP EYKWNMSGTWOAEFTKPTDIEINMGVNGALQMNVLPGKGDWESVRVVKYDQILPV CSILLEYDIYIPDVEGLTRLRPYAVLNPGLWIKGLDMNNTSIDSELVFDGKKYKHF HVRIEFDKTPGVNELHIGVVDHLEVDGFIDNVRLYKKS"	
BASE COUNT	598 a 394 c 580 g 429 t	
ORIGIN		
Query Match	5.4%; Score 124.4; DB 1; Length 2001;	
Best Local Similarity	53.0%; Prod. No. 1.2e-07;	
Matches	424; Conservative 0; Mismatches 316; Indels 60; Gaps 5;	
Qy	376	CCGGCGCTACACGACGCGCGCGCTTCCAGGCGCTTGCATACGCGATCGGAGCGCG 435
Db	298	CCAGAGGGTACGAACTCAGGACGGTTTGAAGAGCTCCAGCTACGAGTAGCGAAGCA 357
Qy	436	GCGCGCACGCGCTTCGGGTGATTGTCCTCCACCACGACTGGAAGAAATTTGGGGA 495
Db	358	AAAGAACTCGGCATAAAGCTCATAATCGTCTTGTGAACAACCTGGACGACTTCGGTGG 417
Qy	496	ATGGATCAATACGACAAGTGGTACGCGCTTCTTACCACGACAACTTCTACACGCGCCC 555
Db	418	ATGAATCAATACGTGAGATGGTTGGGGGCATCCATCAGCATCTCTTACAGGAACGAG 477
Qy	556	CGGACCCACGCGGTACAGAATTCGGTCAATCATCTTACTGAACGGGTCAACAGCAT 615
Db	478	AAAGTCAAAAGAAATACAAAAGTAGTCTCTTCTCTATAAAGAGGTGAACACCTAC 537
Qy	616	ACCGCGCTGACGTACAGAAGCATCCACGATCTTTGTTGGGAAGCTTCCGCAATGAGCCG 675
Db	538	ACGGGTGTTCTTACAGGGAAGAGCCACCATCATGCGATGGCACTGGCAGACGAGCCC 597
Qy	676	CGCTCGGTAGGAAGGGGCACATTAACAACTCGGGCACGTGCACCTCAGGCGACCATTTGTC 735
Db	598	AGGTGTGAACGAGC-----AAGTCTGGTAACACACTCGTT 633

```
QY 736 AACTGGTTCGATCAATATGCGCGTACGTCACAAAGCATAGACCCCTAACCATATGCTCTCG 795
  ||||| || |||| ||||| || |||| || || |||| ||||| ||||| ||
Db 634 GAATGGGTAGAGAGATGAGTCTTACATAAAGAGTCTGGATCCAAACCCACCTGGTCC 693
QY 796 GTCCGGCAGGAGGGTTCTACA-----TTCCGCTCAACGCCAGGGAAGCGCT 841
  || ||||| ||||| || ||||| ||||| || ||||| ||||| ||||| ||
Db 694 GTGGGAGAGGAGGATTTCTTCAACAACACTACGAAGGCTTCAGACCTTACCGTGGAGAGGCT 753
QY 842 G-GCCATACAACGACCCGTCACCGCGTCGACACAACTGCTCTCTCTCGTCTCAAGAAC 900
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 754 GAGTGGGCTACAGCGATGGTCCCGTGTGTGAGTGAAGAGACTTCTGAGATAGAGCG 813
QY 901 ATTGACTTTGGCAGCTATCACCTGTACCCGAATTAAGTGGGGCAGC-----AC 948
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 814 GTGGATTTTGGTACGTTCCATCTTACCCCTCCCACTGGGGTGTGAGCCCTGAACACTAC 873
QY 949 GCGGACTGGGGAACCAATGGATCAAGGATCATATTGCGTAAGCCGACCGATCGGCAAG 1008
  || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 874 GCACAGTGGGGGCAAGTGGATCAAGATCACATAAAGATCCCAAAAGAGGTGGAAAA 933
QY 1009 CCGACCATTTCTGAAGAATTCGGCTGGCAGACACCGG-----ACCGCGATTCCGTC 1059
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 934 CCGCTGCTTCTGAGAGACTACGGTATTCCTCAAAAGTCCCGGTCACACAGGGTTGCCATT 993
QY 1060 TATCAGACGTGGACCCAGACTGTGCGTACGAACGGTCAAGCAGGCTGGAACCTTCTGGATG 1119
  || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 994 TACAAATTCTGGAACCATCTGGTCTACACCTCGGTGGAACCGTGCCCATGTTCTGGATG 1053
QY 1120 CTCGCTGGCAATGTCACCG 1139
  ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1054 CTCGAGGAATCGTGAAG 1073

RESULT 15
LOCUS TNE17981
DEFINITION Thermotoga neopolitana manB gene.
ACCESSION Y17981
VERSION Y17981.1 GI:6006596
KEYWORDS beta-mannosidase; manB gene.
SOURCE Thermotoga neopolitana.
ORGANISM Bacteria; Thermotogae; Thermotogales; Thermotogaceae;
Thermotoga.
REFERENCE 1 (bases 1 to 2001)
AUTHORS Parker,K.N., Lam,D., Duffaud,G., Snead,M.A., Mathur,E.J. and Kelly,R.M.
TITLE Amino Acid Sequence of beta-mannosidase genes from the hyperthermophilic bacteria Thermotoga maritima and Thermotoga neopolitana
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2001)
AUTHORS Parker,K.N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) K.N. Parker, North Carolina State University, Department of Chemical Eng, Box 7905 NCSU, Raleigh, NC 27695-7905, USA
COMMENT sequence incorrect.
FEATURES
    source
        1..2001
            /organism="Thermotoga neopolitana"
            /strain="5068"
            /db_xref="taxon:2337"
    gene
        1..2001
            /gene="manB"
    CDS
        1..2001
            /gene="manB"
            /codon_start=1
            /transl_table=11
            /product="beta-mannosidase"
            /protein_id="CAB56856.1"
            /db_xref="GI:6006597"
```

```
/db_xref="SPTREMBL:Q9RIK7"
/transl_table="MKLVFSLI1VTLPIVLVFNANSDFKVKNGRFLLNGEEFKFVGSN
NYMHYKSNRMDISVLESAKANGVKVLRWGFLDGESCRDKNYTHIPAPGVFLPEG
TNAQDGFRLDYVAKAKELG1KL1VLVNNDDFGGNNQYVRWEGGLIUDDFYRNEK
IKETKTVSFLINRNTYTGPVREEPTIMAWELANEPRETCDKSNLTLYBWEKMS
AYIKSLDPNHLVAVGDEGFNNYEGFRPYGGAEWAYNCGWGVDMKRLLEITVDGFT
PHLYPSHMGVSPENYAOMGAKWIEDH1K1TAKEGVKPVLEEYGTIPKSAVNRVAYIKL
WNLVYNLGGANGAMFWMLAG1GEGWDRDEKYYPDYDFGR1VNDESEAKL1REYAKL
FSTGDTREDTCMF1TPKDG0E1KKTVKVRGVDFYNTFKG1SVGVENLLFDEIKH
LCGYVYGFEFDTTR1SDGEHEMFLEAHFRGETVKDTR1RVKVNRAVYLAESVDFSRP
BEVKNNWSGTWQAEFKTPD1EWNVEGNGALOMNYYLPGKGDNEEVRVVRKFDQLPV
CEILEYDIY1PDVEGTR1PRYAVLNPQWVK1GLDMNNTSIDSGELV5FNKGGKKYHKF
HYRIEFDKTPGVNHLIGVVG0HLEVDGPF1DNVRLYKSS"
BASE COUNT 598 a 394 c 580 g 429 t
ORIGIN
Query Match 5.4%; Score 124.4; DB 1; Length 2001:
Best Local Similarity 53.0%; Pred. No. 1.2e-07:
Matches 424; Conservative 0; Mismatches 316; Indels 60; Gaps 5:
QY 376 CGCGCGTACAACGACGCGCGCGCTTCAAGGCTTCACTACGCGCGACGCGG 435
  || || || || || || || || || || || || || || || || || || || || || ||
Db 298 CCAGAGGTACGACGCTCAGGACGGTTTGAAGACTCGACTACACGGTAGCGAAGCA 357
QY 436 GCGCGCACGCGCTTCGGGTGATTGTCTCTCCACCAACGACTCGGAAGAATTTGGCGGA 495
  || || || || || || || || || || || || || || || || || || || || || ||
Db 358 AAAGAAGCTGGCATAAAGCTCATAATCGTTCTTGTGAACAACCTGGGACGACTTCGGTGA 417
QY 496 ATGGATCAATACGACAAGTGGTACGCGCTTCTTACCACACAACCTTCTACACGCGCCC 555
  || || || || || || || || || || || || || || || || || || || || || ||
Db 418 ATGAATCAATACGTGAGATGGTTTGGGGGATCCATCAGCATCTCTACAGGAAGAG 477
QY 556 CGGACCCAGCAGCGGTACAAGAATTGGGTCAATCATCTAGTGAACGGGTCAACAGCATTT 615
  || || || || || || || || || || || || || || || || || || || || || ||
Db 478 AAGTCAAGAAGAATACAAAAGTACGTCTCTTCTCATAAACAGGTGAACACCTAC 537
QY 616 ACGCGGTACGTACAGAAGCATGTCACGATCTTGTCTTGGGAACCTTGGCAATGAGCGG 675
  || || || || || || || || || || || || || || || || || || || || || ||
Db 538 ACGGTTCTTCTTACAGGGAAGAGCGCACCATCATGTCATGGGAACCTGGCGAAGAGTCC 597
QY 676 CGCTCGGTAGGAGCGGCACATTACCANCTCGGCACCTGTCACACTAGGCGAATCATTTTC 735
  || || || || || || || || || || || || || || || || || || || || || ||
Db 598 AGGTGTGAACCGGAC-----AAGTCTGTGTAACACACATCGTT 633
QY 736 AACTGGTTCGATCAAAATGCGCGTACGTCAAAAGCATAGACCCCTAACCATATGCTCTCG 795
  ||||| || |||| ||||| || |||| || || |||| ||||| ||||| ||
Db 634 GAATGGGTAGAGAGATGAGTCTTACATAAAGAGTCTGGATCCAAACCCACCTGGTCC 693
QY 796 GTCCGGCAGGAGGGTTCTACA-----TTGGGTCAACGCCAGGGAAGCGCT 841
  || ||||| ||||| || ||||| ||||| || ||||| ||||| ||||| ||
Db 694 GTGGGAGAGGAGGATTTCTTCAACAACACTACGAAGGCTTCAGACCTTACGCTGGAGAGCT 753
QY 842 G-GCCATACAACGACCCGTCGCGAGCGCGTCGACAAACATGCTCTCTCTCGTCTCAAGAAC 900
  || || || || || || || || || || || || || || || || || || || || || ||
Db 754 GAGTGGGCTACAAACGGATGGTCCCGTGTGTGACTGGAAGAGAGACTTCTGGAGATAGAGCG 813
QY 901 ATTGACTTTGGCAGCTATCACCTGTACCCGAATTAAGTGGGGCAGC-----AC 948
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 814 GTGGATTTTGGTACGTTCCATCTTACCCCTCCCACTGGGGTGTGAGCCCTGAACACTAC 873
QY 949 GCGGACTGGGGAACCAATGGATCAAGGATCATATTGCGTAAGCCGACCGATCGGCAAG 1008
  || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 874 GCACAGTGGGGGCAAGTGGATCAAGATCACATAAAGATCCCAAAAGAGGTGGAAAA 933
QY 1009 CCGACCATTTCTGAAGAATTCGGCTGGCAGACACCGG-----ACCGCGATTCCGTC 1059
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 934 CCGCTGCTTCTGGAAGAGTACGGTATTCCTCAAAAGTCCCGGTCACACAGGGTTGCCATT 993
QY 1060 TATCAGACGTGGACCCAGACTGTGCGTACGAACGGTCAAGCAGGCTGGAACCTTCTGGATG 1119
  || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 994 TACAAATTCTGGAACCATCTGGTCTACACCTCGGTGGAACCGTGCCCATGTTCTGGATG 1053
QY 1120 CTCGCTGGGAATGTCACCG 1139
  ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
```

Db 1054 CTCGCAGGAATCGGTGAAGG 1073
||||| || | | | ||

Search completed: November 13, 2002, 10:08:04
Job time : 4057 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2002, 08:58:51 ; Search time 326 Seconds
(without alignments)
15812.344 Million cell updates/sec

Title: US-09-917-378-2

Perfect score: 2289
Sequence: 1 atgggtctagtgcgtgcc.....tcacctgcacgctagtgtga 2289

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	180.4	7.9	3004	17 AAT12337	A. cellulolyticus
2	180.4	7.9	3004	21 AAZ55924	Acidothermus cellu
3	180.4	7.9	3004	24 ABK86729	A. cellulolyticus
4	143.6	5.8	11706	20 AAX55661	DNA sequence encod
5	143.6	5.8	11707	24 AAD28525	Active cellulase p
6	128.4	5.6	6415	20 AAX55662	DNA sequence encod
7	128.4	5.6	6416	24 AAD26526	Active cellulase p
8	128	5.6	2043	18 AAT93688	Thermotoga maritim
9	128.4	5.5	1992	19 AAV36921	Thermotoga maritim

10	126.4	5.5	2043	19 AAT94213	Thermotoga maritim
11	124.8	5.5	2043	19 AAV36917	Thermotoga maritim
12	124.4	5.4	2001	19 AAT94200	Thermotoga neapolit
13	114.8	5.0	2977	12 AAQ13001	Endol gene encodin
14	94.6	4.1	1621	19 AAV07163	Thermomonospora fu
15	90.2	3.9	2029	20 AAX55660	DNA sequence encod
16	90.2	3.9	2029	24 AAD26568	Active cellulase h
17	89	3.9	2600	12 AAO15178	Portion of PABC-1
18	82	3.6	777	21 AAC44444	Arabidopsis thalia
19	82	3.6	784	21 AAC47999	Arabidopsis thalia
20	82	3.6	796	21 AAC45005	Arabidopsis thalia
21	82	3.6	2011	21 AAC52204	Arabidopsis thalia
22	81.6	3.6	4056	22 AAH75060	Nucleotide sequenc
23	78.4	3.4	390	13 AAQ21833	Randomising oligon
24	78.4	3.4	390	14 AAQ36859	PCR primer for 5'
25	78.4	3.4	390	22 AAT76910	Sequence containin
26	78.4	3.4	390	24 AAT17275	Oligo #7 for clon
27	77	3.4	4059	22 AAH75059	Nucleotide sequenc
28	75.8	3.3	2293	19 AAV09659	A. cellulolyticus
29	72.6	3.2	38186	20 AAZ32028	Human MTH1 relate
30	72.6	3.2	38186	22 AAC90085	AC004449 cDNA clon
31	70.6	3.1	2308	24 ABQ39168	Oligonucleotide fo
32	70.6	3.1	2308	24 ABQ39169	Oligonucleotide fo
33	70.6	3.1	6510	22 AAS46648	Tumour suppressor
34	70.6	3.1	6510	24 ABL32395	Human immune syste
35	70.6	3.1	6510	24 AAS61115	Human gene regulat
36	69.8	3.0	1232	22 AAD11124	Human small cell l
37	69.6	3.0	2175	14 AAQ49820	NK-1 cellulase gen
38	69.6	3.0	2365	24 ABQ17066	Oligonucleotide fo
39	69.6	3.0	2365	24 ABQ17067	Oligonucleotide fo
40	69.2	3.0	150	20 AAV64956	Mouse histone H2B
41	68.8	3.0	1314	24 ABK73393	Bacillus lichenifo
42	68.2	3.0	1844	21 AAC41572	Arabidopsis thalia
43	68.2	3.0	1844	21 AAC47098	Arabidopsis thalia
44	68.2	3.0	1847	21 AAC38734	Arabidopsis thalia
45	65.6	2.9	2510	24 ABK53202	Bacillus subtilis

ALIGNMENTS

RESULT 1
AAT12337 standard; DNA: 3004 BP.

XX	AAT12337;				
XX	08-OCT-1996 (first entry)				
XX	A. cellulolyticus E1 endoglucanase gene.				
XX	El endoglucanase; cellulase; cellulose; saccharification; ethanol:				
XX	ss.				
OS	Acidothermus cellulolyticus.				
XX	Key	Location/Qualifiers			
FT	misc_difference	200			
FT		/note= a			
FT	misc_signal	710..725			
FT		/tag= b			
FT	RBS	/note= putative upstream regulatory sequence*			
FT		772..779			
FT		/tag= c			
FT	repeat_unit	781..790			
FT		/tag= d			
FT	repeat_type= DIRECT				
FT		795..804			
FT		/tag= e			
FT		/rpt_type= DIRECT			

OY 2032 GGCTTTACGGCAGCGGTGACGGTACGAATACCGGAGCGCGGCGACGCGCGGTGGACG 2091
 II IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 2255 GGCTTCACGGTACCGGTGCGCGTGACAAATTCGGATCCGTCCGGACCAAGACATGGACG 2314
 OY 2092 GTGCGCGTGGTTCGTTGGTGGGAATACAGCGGTACAGAACTACTGGACACTGCGTTGACG 2151
 II IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 2315 GTCACTGGACATTCGCGCGGAATACAGACGATTCACCAATTCGTGGAAATCGACGGGTCA 2374
 OY 2152 CAATCAGGTGCATTCGCTGACGGCGACGAACCTCAGCTACAAACCTGATCCAAACCGGT 2211
 II IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 2375 CAGAAGCGTCAGTCGCTGACGGCTCGAATATGAGTTATACAAACCTGATTCAGCTGGT 2434
 OY 2212 CAGTCGACCACTTCGGATTCAACGGAAGTTACTAGGAACAACACCCACCTACACTC 2271
 II IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 2435 CAGAACACCACTTCGGATTCCAGGCGAGCTATACCGGAACGAACGCGGACCGACAGTC 2494
 OY 2272 ACCTGCACCGCTAGTTCA 2289
 II IIII IIII IIII IIII
 Db 2495 GCCTGCGCAGCAAGTTAA 2512

RESULT 4

AAX55661
 ID AAX55661 standard: DNA; 11706 BP.

AC AAX55661;

XX 30-JUL-1999 (first entry)

XX DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KW cotton-containing fabric; stonewashing; ss.

XX Unidentified.

XX EP921188-A2.

XX 09-JUN-1999.

XX 15-SEP-1998; 98EP-0810919.

XX 19-SEP-1997; 97US-0932571.

XX (CLRN) CLARIANT FINANCE BVI LTD.

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;

PI Gibbs MD, Morgan H, Williams DP;

XX WPI: 1999-315403/27.

DR P-PSDB: AAY13492.

XX New truncated cellulase proteins, useful in detergents and for
 PT producing 'stonewashed' denim

PS Disclosure: Page 20-23; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful

CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

XX Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 other:

SQ Query Match 5.8%; Score 133.6; DB 20; Length 11706;
 Best Local Similarity 52.4%; Pred. No. 2.5e-16;
 Matches 395; Conservative 0; Mismatches 344; Indels 15; Gaps 4;

OY 1220 CAATCAGTACCGGCACATCCCTCCGCCGTGCGGAGCTCGAGTCCATCTCTGTCGCGGT 1279
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 3883 CAACACACCTGCACGCACATCAGCCGCCGACACCGAGCCCAACAGTCACAGCAACCCGCA 3942
 OY 1280 CTCGTCGCGGTCTCCGTGCGCGTCTCCGTGCGGCTCTCCGTGCGGCTCTCTGTCGCGCA 1339
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 3943 CTCCAACGCGCGCGACAGTACGGTTACTGTGACTCCGACACCCACCAACACCGCA 4002
 OY 1340 GCCCGTCTCCGTCTGTCGCCGTGTTCGG---GTGGGTGAAGTGCAGTACAGAACA 1396
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4003 CGCGACACCGACAGGACACCTGGCACGGGAAGTGGTTGAAGGTACTATACAAGAACA 4062
 OY 1397 ATGATTCCGCGCGGTGATTAACACAGATCAACCGGCTCTCCAGTTGGTGAATACCGGT 1456
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4063 ATGAGACAAGTCGAGCACAAGTTCTATAAGCGCGTCTTAAAGATAGTCAATGGAGCA 4122
 OY 1457 CGTCGTCGGTGGATTTGTCACCGGTGACGTCGGGTACTGTTGTTACCCCGGATGGTGGT 1516
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4123 GCAGCAGTGTGATCTTAGCAGGGTTAAGATAAGATACTGGTACACAGTGGATGGTG--- 4179
 OY 1517 CGTCGACACTGGTGTACAACCTGCACTGGCGGCGGTGGGTGTGGGAATATCCGCCCT 1576
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4180 ACAAGCACACAGTGGGTATGTGACTGGCCACAGATAGGGGACAGCAATGTGACATTC 4239
 OY 1577 CGTTGCGCTCGGTGAACCCCGGACCGGACGCGGACACTACCTGCAGTTGTTCGTTC 1636
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4240 ATTTGTGAAGCTGAGCAGCGGAGTCAAGTGGAGCGGATTTATTACTTGGAGGTAGGATTT 4299
 OY 1637 CTGCTGGAAC-----GTTGGCGCTGGTGGGTGCGAGGGTGAGATTCAAAACCGGGTGA 1690
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4300 GCAGTGGAGTGGGCGAGTTGCACGCTGGTAAGACACACAGGGGATATACAGGTAAAGTTTA 4359
 OY 1691 ATAAGAGTACTCGTCAACTTTTGTATGAGACCAATCTACTCTGTATGGACGAACACCG 1750
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4360 ACAAGAATGACTGGAGCAATTTACAATCAGCAGACGACTGGTCATGCTTCAGAGCATGA 4419
 OY 1751 CCTTCCAGGATTCGA---CGAAGTGACGCTGTATGTCATGCGCGGCTGGTGTGGGGA 1807
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4420 CGAATTATGGAGAGAAATGCGAAGGTAAACGCTGTATGTAGATGCTGTCTGTTCTGTTGGGGC 4479
 OY 1808 CTGAACGCTCCGGCAGCAGCGCCACACCGCCACACCGCCCGCCCAACCCCGTCCCGCA 1867
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4480 AGGAGCGGGGAGGAGGACACCTGCACCGCAACAGCAGCAGCAACCAACCCCACTCCGA 4539
 OY 1868 GCCCGAGCGCGACCCCAAGCCCGAGCTCTCTCCCATCCCGTCCCGAGGCCCGACGCCCA 1927
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4540 CAGCAACAGCAACACCGCGCCGACAGCAACGCCACGCTCTACACCGACACGACAGCA 4599
 OY 1928 GCCCTACGCGGTCCCGTCCCGAGCGCGCTGCGCC 1961
 IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 Db 4600 CACCAACCCCAATPACCAACACCCCAACCGCTCC 4633

RESULT 5

AAD26525

ID AAD26525 standard: DNA; 11707 BP.

AC AAD26525;

XX 26-MAR-2002 (first entry)

XX Active cellulase protein, celB gene.

PT producing 'stonewashed' denim
PS Disclosure: Page 24-25; 65pp; English.
XX
CC The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkalophilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAV13492; Cel E1 extends
CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAV13493; Cel E3/B5 is shown in AAV13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or
CC remove staining, backstaining or graying, for use on cellulosic
CC materials including cotton-containing fabrics. They are especially useful
CC for preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using
CC non-truncated cellulase compositions.
XX
SQ Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 36 other;
Query Match 5.6%; Score 128.4; DB 20; Length 6415;
Best Local Similarity 52.6%; Pred. No. 2.3e-15;
Matches 379; Conservative 0; Mismatches 326; Indels 15; Gaps 4;

QY 1243 CCGCGCTCGCGAGCTCAGTCCATCTCTCGTCCGCGTCTCGTCCGCGTCTCGTCCGCGG 1302
DB 2532 CCGGTCGGGCAACATCCACACCTCCACCAACATCAACACCAACCAACCCCGACCCCA 2591
QY 1303 TCTCTCGTCCGCGCTCTCGTCCGCGAGCCGCTCTCGTCCGCGTCCGCGG 1362
DB 2592 ACACCAACAGTGACAGCAACCGCGCGACTCTACACCGACACCGCGGTCACT 2651
QY 1363 GTGTCGG---GTGGGGTGAAGTGTGAGTACAGACAAATGATTCGGCGCCGGGTGATAAC 1419
DB 2652 GGTACCGGAAGTGTGTGAAGTACTGTACAGAACAATGACAGAAGTCCGAGCAGGT 2711
QY 1420 CAGATCAACCGGGTCTCCAGTGTGTGAATACGGGGTCTCGTCCGTTGATTTGTCGAGG 1479
DB 2712 TCTATAAGCGCGTGTGTTAAGTAGTGAATGAGGCGACGAGTGTGATCTTAGCAGG 2771
QY 1480 GTACCGGTCCGCTACTGTTTACCCCGGATGTTGGTCTCGACACTGTTGTACAACCTGT 1539
DB 2772 GTTAAGATAAGATACTGTGTACACAGTGATGTGTG---ACAAGCCACAGTCCGGTATGT 2828
QY 1540 GACTGGCGCGGATGGGTGTGGGAATATCCCGGCTCGTTGGGTCGGTGAACCCGGCG 1599
DB 2829 CACTGGGACAGATAGGGGCAAGCAATGTGACATTTCAATTTTGTGAAGCTTAGCAGCGGA 2888
QY 1600 ACGCGCGCGGCGACACCTACTCGAGTGTCTGTTCACTGTGTGAAC-----GTTGGCC 1653
DB 2889 GTGAGTGAGCGGATTATTACTGGAGTAGGATTAGCAGTGGAGCTGGGCGATTGCGAG 2948
QY 1654 CTGCTGGTCCGACGGGTGAGATTCAAAACCGGTTGAATAAGAGTCACTGTGCGAACTTT 1713
DB 2949 CTTGTAAGGACACAGGGGATATACAGGTTTAAACAGATGATGAGCAATATAC 3008
QY 1714 GATGAGCAATGACTACTCTGATGGGACAGACACCGCTTCCAGGATTGGA---CGAAG 1770
DB 3009 AATGAGGACAGCATGCTGTGATGTTGTCAGAGCATGACGAATATTGAGAGAAATCCGAAG 3068
QY 1771 GTGAGGTGTATGTAATGTGGCGCTGTGTGGGGACTGAACGCTCGGCGACACGCC 1830
DB 3069 GTGAGCTGTATGATGTGTGTTCTGTGTGGGGAGGAGCCGGGAGGAGCGACACT 3128
QY 1831 AGCCCCACACCCAGCCAGCCAGCCAGCCGCTCCCGAGCCGAGCCGACCCCAAGCC 1890
DB 3129 GCACCGAAGACAGCAACCAACCAAGCCCACTCCGACAGCAACCCCAACCACTACACT 3188

QY 1891 AGCTCTCCCATCCCGTCCCGAGCCCGAGCCCGAGCCCTACGCGTCCCGTCCCGTCCCGG 1950
DB 3189 ACACCGACCCCGACCGACAGTGAAGCAAGCCCAACACCGACCGACCGATCACCG 3248

RESULT 7
AAD26526
ID AAD26526 standard; DNA; 6416 BP.
XX
AC AAD26526;
XX
DT 26-MAR-2002 (first entry)
XX
DE Active cellulase protein, celE gene.
XX
KW Active cellulase protein; alkalophilic; textile processing; proteinase;
KW detergent additive; stonewashed appearance; cotton-containing denim;
KW CelB5; thermophilic; commercial detergent; celE gene; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 634..5889
FT /*tag= a
FT /product= "CelE protein"
FT misc_feature 748..2076
FT /*tag= b
FT /product= "DNA encoding E1 protein"
FT misc_feature 748..2538
FT /*tag= c
FT /product= "DNA encoding E1/2 protein"
XX
PN US6294366-B1.
XX
XX 25-SEP-2001.
XX
PF 19-AUG-1998; 98US-0136574.
XX
PR 19-SEP-1997; 97US-0932571.
XX
PA (CLRN) CLARIANT FINANCE BVI LTD.
XX
PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
PI Morgan H, Williams DP;
XX
DR WPI: 2002-081780/11.
DR P-PSDB; AAE16324.
XX
PT New cellulase active protein, useful in textile processing or
PT commercial detergents, e.g. for improving the feel or appearance of
PT cotton-containing fabrics, is stable under conditions of alkaline pH
PT and elevated temperatures -
XX
PS Disclosure: Column 37-44; 61pp; English.
XX
CC The present invention relates to a cellulase active protein, which is
CC substantially free of proteinases of native thermophilic and
CC alkalophilic origin, where the cellulase active protein consists of the
CC CelB5 amino acid sequence. The cellulase active protein is useful for
CC treating cellulosic materials including cotton-containing fabrics, as
CC detergent additives. The cellulase active protein is also useful for
CC improving the feel and/or appearance of cotton-containing fabrics, for
CC removing surface fibers from cotton-containing knits or for imparting
CC stonewashed appearance to cotton-containing denims. The present proteins
CC are stable under condition of alkaline pH and elevated temperatures,
CC thus suitable for textile processing and in commercial detergents.
CC The present sequence is celE gene.
XX
SQ Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 36 other;
Query Match 5.6%; Score 128.4; DB 24; Length 6416;
Best Local Similarity 52.6%; Pred. No. 2.3e-15;
Matches 379; Conservative 0; Mismatches 326; Indels 15; Gaps 4;

```
QY 1243 CCGCGTCCGCGAGCTCGAGTCCATCTCGTCCGCGCTCTCGTCCGCGCTCTCGTCCGCG 1302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2533 CCGGTCGGCAGACATCCACATCGACCAACATCAACACCAACACCCCGACCCCA 2592
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1303 TCTCCGTCGGCGTCTCGTCCGCGCTCTCGTCCGCGAGCGCTCTCGTCCGTCGTCGCG 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2593 ACACCAACAGTGACAGCAACGCCGCCGACTCTCTACACGACACCGCGGTCACCT 2652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1363 GTCTCGG---GTGGGTGAAGTGCAGTACAAGAACATGATTCGCGCGCGGTGATAAC 1419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2653 GTTACGGGAAGTGTGTGAAGTACTGTACAAGAACATGAGACAACTGCGAGCACAGGT 2712
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1420 CAGATCAAAACCGGCTCTCCAGTTGGTGAATACCGGGTCTCGTCCGTCGATTTGTCGACG 1479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2713 TCTATAAGCGCGTGTGTTAAAGATAGTCAATGAGGCGACGACGCTGTTGATCTTAGCAGG 2772
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1480 GTACGGTGGGTACTGTTTACCGCGGATGTTGGTCTCGACACATGTTGTACAACTGT 1539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2773 GTTAAGATAAGATACTGGTACACAGTGGATGGTG---ACAAGCCACAGAGTGGCGTATGT 2829
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1540 GACTGGCGCGGATGGGTGTGGGAATATCCGCCCTCTGTCGCTCGGTGAACCCCGCG 1599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2830 GACTGGGCACAGATAGGGCAAGCAATGTGACATTTCAATTTGTGAAGCTTAGCAGCGGA 2889
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1600 AGCCGACGCGGACACCTACTCGTCACTGTGTTGTTCACTGTGGAAC-----GTTGGCC 1653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2890 GTGAGTGGAGCGGATTTATACCTGGAGGTAGGATTTAGCAGTGGAGCTGGGCGACTTGCAG 2949
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1654 GCTGGTGGTCTGACGGGTGAGATTCAAACCGGGTGAATAGAGTGACTGTCGAACTTT 1713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2950 CTTGGTAAGACACAGGGGATATACAGTAAAGTTTAAACGAAGTACTGGAGCAATTAC 3009
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1714 GATGACACCAATCACTCTGATGGACGAACACCCCTCTCCAGGATTTGA---CGAAG 1770
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3010 AATCAGCAGACGACTGCTATGCTTCAGAGCATGACGAATTTATGAGAGAAATGCGAG 3069
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1771 GTACGGTGTATGTCATATGCGCGCTGTGTGGGGACTGAACCGTCCGGACACGACCCC 1830
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3070 GTACCGCTGTATGATAGTGTGTTCTGTTATGGGGGAGGAGCGGAGGACACCT 3129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1831 AGCCCCACACCCAGCCCCAGCCCAACCCCGTCCCGAGCGCGGACCCCAAGCCCC 1890
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3130 GCACCGACACGACGACCAACCCCACTCCGACAGCAACCCCAACACCTTACACCT 3189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1891 AGTCTCTCCCATCCCGTCCCGAGCCCCAGCCCCAGCCCTACGCGCTCCCGCTCGCG 1950
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3190 ACACCGACCCGACACGACGAGTGAGTGCAACGCGCAACACCGGACCGGACGCGCATCAG 3249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 8
AAT93688
ID AAT93688 standard; DNA; 2043 BP.
XX
AC AAT93688;
XX
DT 12-MAR-1998 (first entry)
XX
DE Thermotoga maritima beta-mannanase encoding DNA.
XX
KW Glycosidase; thermostable; textile; food processing; pharmaceutical;
KW detergent; baking; industry; Thermococcus; Staphylothermus;
KW Pyrococcus; glucose; soluble oligosaccharide; beta-mannanase; ss.
XX
OS Thermotoga maritima.
XX
FH Key Location/Qualifiers
FT 1..2043
FT /*tag= a
FT /product= Beta_mannanase
XX
PN W09725417-A1.
XX
```

```
PD 17-JUL-1997.
XX
PF 10-JAN-1997; 97WO-US000092.
XX
PR 13-SEP-1996; 96US-0712612.
PR 11-JAN-1996; 96US-0583787.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
XX
DR WPI; 1997-372858/34.
DR P-PSDB; AAW34564.
XX
XX New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
XX and Pyrococcus, used in the textile, food processing,
XX pharmaceutical, detergent and baking industries
XX
XX Claim 4; Fig 11; 82pp; English.
XX
CC The present sequence encodes beta-mannanase isolated from Thermotoga
CC maritima. The enzyme or its encoding nucleic acid sequence is
CC used for generating glucose from soluble oligosaccharides. The enzyme
CC can be used in the food processing, pharmaceutical, textile, detergent
CC and baking industries. The enzyme is also used to treat lactose
CC intolerance, as a diagnostic reporter molecule, in corn wet milling or
CC in the fruit juice industry. The enzymes can be used to hydrolyse guar
CC gum to remove non-reducing terminal mannose residues. The nucleic acids
CC encoding the enzyme may be used to generate probes to identify similar
CC sequences.
XX
SQ Sequence 2043 BP; 609 A; 423 C; 587 G; 424 T; 0 other;
XX
Query Match 5.6%; Score 128; DB 18; Length 2043;
Best Local Similarity 49.4%; Pred. No. 2.6e-15;
Matches 514; Conservative 0; Mismatches 475; Indels 51; Gaps 5;
QY 121 TTCGTACCGCATCCGGCGTCACTTCGTTCTGAACGCCCTTCCCTATCTGTTACGGGGA 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 TTCGTGAAGTGGAAACGGAAATTCGCTCTGAACGGAAAGAAATTCAGATTCATTGGA 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 AGCAACAACTATTACCTAGCTATCAGTCGCGACGCCGACGTCGATGAGCTGTGGCCAA 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 AGCAACAACTACTACATGCACCTACAAGACCAACGGAATGATAGACAGTGTCTGACAGT 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 GCTCAACGAGTAACTTCTCTCATCGGACCTGGGCTTTCATCGACATCGGCTCTCTT 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 CCCAGACATGGGTATAAGGTCTCAGAACTCTGGGCTTTCCTCGAGGGGGAGAGTTAC 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GACGGCTCCGTGCCCAACATCGATGGCAACAAAGAACGCCCTTCTACTTTTCACTACTGGGAC 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 TGCAGACACAAGAACACCTACATCATCTCAGCCCGGTGTTTCGGGGTCCCAAGAGA 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 CCGTCGACCGCGCTCCGGGTACACGCGGCGGCGGCGCTTGCAGGCGCTTGACTAC 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 ATATCGAACGCC-----CAGACGGGTTCGAAAGACTCGACTAC 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 CGCATCCGAGCGCGCGCGCATCGGCTTCGGGTGATTTGCTCTCCCAACGACTGG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 ACATTCGGAAGCGGAAGAACTCGGTATAAATCTGTCATTTGTTGTAACAACTGG 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 481 AAAGAATTTGGGGAATGGATCAATAGCACAAGTGGTACGCGCTTCCCTTACGACGACAAC 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 439 GACGACTTCGTGGAATGAACCACTACGTAGGTGGTTTGGAGGAACCCATCAGCACT 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 541 TTCTACACCGACCCCGGACCGCGGTGACAAAGAAATTTGGGTCAATCATCTACTGAAC 600
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 499 TTCTACAGAGATGACAAGATCAAAAGAGAGTACAAAAGTACGTCCTCTCTCTGTAAC 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 601 CGGTCACACGACTTACCGGCGTGACCTACAAGAACCATCAACGATCTTTTGTCTGGAA 660
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 559 CATGTCATACACGGGAGTTCTTTACAGGGAAGAGCCCACTCATCGCTGGGAG 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


QY 841 TGGCCATACAAACGACCCGTCGCCAGCGGTCGACAAACAATGCTCTTCTCGTGTCAAGAAC 900
DB 742 TGGGCCCTACAACCGCTGGTC---CGGTGTGACTGCAAGAAGCTCCTTTTCGATAGACGC 798
QY 901 ATTGACTTTGGCACGTATCACCTGTACCCGAATTACTTGGG-----CCAGAAC 948
DB 799 GTGGACTTCGGCACGTTCCACCTCTATCCGTCCTCCACTGGGGTGTCACTCCAGAGAATCTAT 858
QY 949 GCGGACTTGGGAACGCAATCGATCAAGGATCATATTGCCGAATGCCGACGCGATCGGCAAG 1008
DB 859 GCCAGTGGGGAGCGAAGTGGATAGAGACCACATAAAGATCCCAAGAGATCGGAAAA 918
QY 1009 CCGACCATTTCTCGAAGAAATTCGG-----CTGCGACACACCGGACCGGATTCGGTC 1059
DB 919 CCGCTTTGTCGAAGAATATGGAATTCGAAGAGTCCGCCAGTTAACACAGCGGCCATC 978
QY 1060 TATCAGACGTGGACCCAGACTGTGCGCTACGAAACGGTGAACGCGCTGGAACTTCTCGATG 1119
DB 979 TACAGACTCTGNAACGATCTGTCTACGATCTCGGTGGAGATCGGAGCGGATGTTCTGGATG 1038
QY 1120 CTGCTGGGAATGTCAACGG 1139
DB 1039 CTCGCGGAATCGGGAAGC 1058

RESULT 10
AAT94213
ID AAT94213 standard; DNA; 2043 BP.
AC AAT94213;
XX
DT 21-MAY-1998 (first entry)
DE Thermotoga maritima endoglucanase DNA.
DE
KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase; ss.
XX
OS Thermotoga maritima MSB8 (clone 6GP2).
XX
PN WO9744361-A1.
XX
PD 27-NOV-1997.
XX
PF 22-MAY-1997; 97WO-US08793.
XX
PR 22-MAY-1996; 96US-0651572.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Lam DE, Mathur EJ;
XX
XX WPI; 1998-018435/02.
DR P-PSDB; AAW35005.
XX
PT Endoglucanase(s), preferably form archaeal bacterium, AEP11 la -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose
PS Claim 3; Fig 10; 164pp; English.
XX
XX
CC This DNA sequence from Thermotoga maritima MSB8 (clone 6GP2)
CC encodes an endoglucanase (see AAW35005) that is able to degrade
CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic
CC bonds in cellulose, and which shows homology to the thermostable
CC endoglucanase (see AAW34985) of archaeobacterium hydrothermal vent
CC isolate AEP11a. The DNA can be used in the recombinant production
CC of the endoglucanase and as a probe to identify similar sequences.
CC 24 Endoglucanase polynucleotides (see AAT94193-216) are claimed.
CC These can be incorporated into plasmid or virus-derived vectors for
CC use in a claimed method of producing enzymes in transformed host

CC cells. The claimed endoglucanases (see AAW34985-W35008) can be used
CC to degrade cellulose for the conversion of plant biomass into fuels
CC and chemicals, for use in detergents, textiles, animal feed, waste
CC treatment, and in the fruit juice and brewing industries for the
CC clarification and extraction of juices.
XX
SQ Sequence 2043 BP; 609 A; 417 C; 592 G; 425 T; 0 other;

Query Match 5.5%; Score 126.4; DB 19; Length 2043;
Best Local Similarity 49.3%; Pred. No 5.1e-15;
Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;

QY 121 TTCGTCAACCCATCCCGCGGTCTGTTCTCAACCGGCTTCCCTATCGTTACGGGGA 180
DB 100 TTCGTGAACTGCAAAACGGAATAATTCGCTCTCAACGGAAGAATTCAGATTCATTGA 159
QY 181 ACGAACAATATTACCTCAGCTATCATGTCGACGCGCCAGCTGATGACGTGTGGCCAA 240
DB 160 AGCAACAATCTACTACATGCACTACAAGACGACGGAATGATAGACAGTGTCTGGAG 219
QY 241 GCTCAACGCGATGAATCTTTCTGTCATCCGACCTGGGGTTTTCATCGACATCGGCTCT 300
DB 220 GCCAGAGACATGGGTATAAGGTCTCTCAGAACTCTGGGGTTTCTCGACGGGAGAGT 279
QY 301 GACGGTCTCGTCCCAATTCGATGGCAACAAGACGGCTTCTACTTTTCAGTACTGGAC 360
DB 280 TGCAGAGACAAGAACACCTACATGCTGCTGAGCCCGGTCTTTTCGGGGTCCCAAGAG 339
QY 361 CCGTCGACCGCGCTCCGGGCTACACGACGCGGCGGCTTGCAGGCTTGCATAC 420
DB 340 ATATCGAACGCC-----CAGACGGTTTTCGAAAGACTCGCATAC 378
QY 421 CGCATCCGACGCGCGCGCGCTTCCGGTGTATCTGCTCTCTCAACACGACTGG 480
DB 379 ACAGTTGCGAAACGGAAGAACTCGGTATATAAACTTCTCATTTCTTCTGAACAAC 438
QY 481 AAGAATTTGGGGGAATGGATCAATACGCAAGTGGTACGGCTTCTTACCACGACAA 540
DB 439 GACGACTTCGGTGAATGAACCACTAGTACGTGAGGTGTTTGGAGGAACCCATCAG 498
QY 541 TTCTACACGACCCCGGACCGGCTACAAGAAATTTGGTCAATCATCTACTCAAC 600
DB 499 TTCTACAGAGATGAGAAGATCAAGAAGAGATCAAAAAGTACGTCTCTTCTCGTA 558
QY 601 CGGCTCAACAGCATTTACCGCGGTGAGTACAAGAAGATCCAAACGATCTTTGCTGG 660
DB 559 CATGCTCAATACCTACACGGGAGTTCCTTACAGGAAGAGCCACCATCATCGGCT 618
QY 661 CTTGCCAATGACCGCGCTGCGTAGGAAGCGGCACATTTACCAACCTCGGGCAGCT 720
DB 619 CTTGCAACGCAACCGCGCTGTGAGACGGAATAATCGGGGAACACGCTCTGTAGT 678
QY 721 CAGGCGACCATTTGTCACCTGGGTGCTCAATGCTCGGCTACGTCAAAAACATAGAC 780
DB 679 AAGGAGA-----TGAGTCTCTACATAAAGAGTCTCGATCCCAACACCTCGTGG 732
QY 781 AACCATATGCTCTCGGTCGCGCACGAAGGTTTCTACATTTGGGTCAACGCAAGGCG 840
DB 733 GGGGACGAAGGATTTCTCAGCAACTACGAAGATTCAAACTTACGGTGAGAAAGCCG 792
QY 841 TGCCCATACAGACCCCGTCCGACGCGGTGCGACAACATGCTCTTCTCCGTGTCAAGA 900
DB 793 TGGGCTTACAACGGCTGGTC---CGGTGTGACTGGAAGAAGCTCTTTTCGATAGAG 849
QY 901 ATTGACTTTGGGACGCTATCACCTGTACCCGAATTTACTGGG-----CCAGAAC 948
DB 850 GTGGACTTCGGGACGCTTCCACCTCTATCGTCCCATCGGGGTGTGAGTCAGTCCAG 909
QY 949 GCGGACTGGGGAACGAATGGATCAAGGATCATATTGCCGAATGCCGACGCGATCGC 1008
DB 910 GCCCAGTGGGGAGCGAAGTGGATAGAGACCACATAAAGATCGCAAAAGAGATCGG 969
QY 1009 CCGACCATTTCTCGAAGAATTCGG-----CTGGCAGACACCGGACCGGATTCGG 1059

Db	970	CCCGTTGTTCTTGAGAAATATGGAATTC	AAAGATGCCGACGTTAC	AGAACGGCCATC	1029
Qy	1060	TATCAGACCTGGACCAGACTCTGCGTACG	AACTGAAACAGGCTTGA	ACTTCTCGATG	1119
Db	1030	TACAGACTCTGGNACGATCTGCTACGAT	TCGGTGGAGATGAGCGATGTTCTCGATG		1089
Qy	1120	CTCGCTGGGAATGTC	AAAGG		1139
Db	1090	CTCGCGGAATCGGGAAGG			1109

RESULT 11

ID AAV36917 standard; DNA: 2043 BP.

AA AAV36917:

21-DEC-1998 (first entry)

DE *Thermotoga maritima* 6CP2 beta-mannanase gene coding region.

AA Glycosylase; 6GP2; thermostable enzyme; oligosaccharide; glucose;
KW baking; textile; detergent; beta-mannanase; ss.

OS *Thermotoga maritima* strain 6GP2.

PN WO9824799-A1.

11-JUN-1998.

XX PF 08-DEC-1997:

XX
PR 10-OCT-1997. 97JIS-0949026

PR 06-DEC-1996; 96US-0056916.
yy

PA (DIVE-) DIVERSA CORP.

PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;

DR WPI; 1998-362407/31.

XX
F 300B, JAN 4 2000.

PT Glycosidase enzymes from organisms of the genera *Staphylothermus*,
PT *Pyrococcus* and *Thermococcus* - for deriving sugar from
PT oligosaccharides, useful in the e.g. food processing, textile or
PT baking industries

PS Claim 1: Flg 11a-d: 92pp: English.

This isolated polynucleotide comprises a coding region for beta-mannanase 6GP2 (see AA049868) from a *Thermotoga maritima* clone (6GP2) that grows optimally at 85 degC in high salt medium. The sequence shows 37% nucleic acid identity to a mannanase of *Aspergillus aculeatus*. The invention provides 18 polynucleotides (see AA36907-24) coding for thermostable glycosidases (see AA049858-75) having glucosidase, alpha-galactosidase, beta-galactosidase, beta-mannosidase, alpha-mannosidase, endoglucanase or pullulanase activity. Vectors and host cells are also claimed. A method is provided for producing the enzymes by recombinant techniques. A claimed method for generating glucose from soluble cell oligosaccharides comprises contacting a sample (selected from dairy products, fruit juice, detergent, textile, guar gum, animal feed, plant biomass or waste product) containing oligosaccharides (selected from maltose, cellobiose, lactose, sucrose, raffinose, stachyose, verbascose, cellulose, starch, amylose, glycogen, disaccharides, polysaccharides and pullulan) with one of the claimed glycosidases such that glucose is produced.

SQ Sequence 2043 BP; 610 A; 419 C; 589 G; 425 T; 0 other;

Query Match

BEST LOCAL SIMILARITY 49.28; Pred. No. 1e-14;

Matches	512:	Conservative	0:	Mismatches	477:	Indels	51:	Gaps
Qy	121	TTCTGTCACCGCATCGCGCGGT	CAGTTCTTCTGTAACCGCCTTCCCTATCTGTTACGGGGA	180				
Db	100	TTCTGTAAGTGGAAACCGAAATTCGCTCTGAACGGAAAGAAATTCAGATTCTATCGA	159					
Qy	181	ACGAACAACATATACCTCAGGTATCAGTTCGACCGCGACGTCGATGACGTGTGGCCAA	240					
Db	160	ACCAACAACATCTACTATCATGTCATACGAAGACAAACGGAATGATAGACAGTGTCTCGAGAGT	219					
Qy	241	GCTCAAGCATGAATCTTTCTGTCTATCGGACCTCGGGCTTTTCATCGACATCGGCTCTCTT	300					
Db	220	GCCAGAGACATGGGTATAAGGTCCTCAAGATCTGGGGTTTTCTCGCGGGAGAGTTTAC	279					
Qy	301	GACGGCTCCGTGCCACAAATCGATGGCAACAAGAACGGCTTCTACTTTTCAGTACTGGGAC	360					
Db	280	TGCAGACACAAGAACACCTACATGCATCCTGAGCCCGGTGTTTCGGGGTCCCAAGGA	339					
Qy	361	CGGTGGACCGGGGTTCGGGTGTACAACGACGGGCGACGGCTTGTCAAGGCTTTGACTTAC	420					
Db	340	ATATCGAACGGC-----	378					
Qy	421	CGGTCGGGACGGCGCGCCACGGCTTCGGGGTGATTGTTCGTCCTCACCAACGACTGG	480					
Db	379	ACAGTTGCGGAAGGGAAGAAGTCCGGTATAAACTTGTTCATTTGTCTGTGAACAACACTGG	438					
Qy	481	AAAGAATTTGGGGAAATGGATCAATACCACAAGTGGTACGGCTTCTCCCTACCAACGACAA	540					
Db	439	GACGATTCGGTGGATGAACCAAGTACGTGAGTGGTGTGGAGGACCCATCAGACGAT	498					
Qy	541	TTCTACACCGACCCCGGACCGGTACAGGATTCGAAGAAATGGGTCAATCATCTACTCAAC	600					
Db	499	TTCTACAGAGATGAGAAGATCAAAAGACAGTACAAAAGTACGTCTCCTTTCTCTGTAAC	558					
Qy	601	CGGTCAACACGATTTACCGGGGTGACGTACAAAGAACGATCCAAAGCATTTTGGTTCGGAA	660					
Db	559	CATGTCAATACCTACACGGGAGTTCCCTTTACAGGAAGACGCCACCATCGGCTGGGAG	618					
Qy	661	CTTGCAATGAGCGCGCTGCGTAGGAAGCGGCACATTTACCAACCTCGGGACGTCGCACT	720					
Db	619	CTTGCAACGACCGCCCTGTGACACGGACAATTCGGGGAACAGCTCGTTGAGTGGGTG	678					
Qy	721	CAGCGACCATTTGTCACTGGGTGCGTCAATGTGCGGTACGTCAAAAGCATAGACCTT	780					
Db	679	AAGGAGA-----TGAGCTCTACATAAAGAGTCTGGATCCCAACCACTCTGTCGTG	732					
Qy	781	AACCATATGTTCTCGGTGCGGACGAAGGTTCATATTGGGTCAACCCAGGGAAGCGGC	840					
Db	733	GGGACGAAGAGATTCTTCAGCAACTACGAAGGATTCAAACCTTACGGTGGAGAAGCCGAG	792					
Qy	841	TGGCCATACAAGACCGCTCGCGCGGTCCACAAACATGCTCTCTCCGTCTCAAGAAC	900					
Db	793	TGGGCTACAACGGCTGGTC---CGGTGTGACTGGAAAGAGTCCCTTTTCATAGACAGC	849					
Qy	901	ATTGACTTTGGCAGTATCAGCTGTACCGCAATTTACTGGGG-----CCAGAAC	948					
Db	850	GTGGACTTCGGCAGCTTCCACCTCTATCCGTCCCACTGGGTGTGAGTCCAGACAACTAT	909					
Qy	949	GCGGACTGGGAACCGCAATGGATCAAGATCATATTGGGAATGCCGAGCGATCGGCAAG	1008					
Db	910	GCCAGTGGGAGCGAGTGGATAGAAGACCAATAAAGATCGCAAAAGATCGGAA	969					
Qy	1009	CCGACCATTCGGAAGAAATTCGG-----CTGGGACACACCGGACCGGATTCCTC	1059					
Db	970	CCCTGTGTTGGGAAGAAATGGAAATTTCCAAGAGTCCGGCAGTTTACAGAACCGCCATC	1039					
Qy	1060	TATCAGACGTGGACCCAGACTGTGCGTACGAACCGTGAAGCGGCTGGAATCTCTGGAT	1119					
Db	1030	TACAGACTCTGGAACGATCTGCTCTACGATCTCGTGGAGATGGAGCGATGTTCTGGATG	1089					
Qy	1120	CTCGCTGGGAATGCAACGG	1139					
Db	1090	CTCGCGGAATTCGGAGAAGG	1109					

RESULT 12
AAT94200
ID AAT94200 standard; DNA; 2001 BP.
XX
AC AAT94200;
XX
DT 21-MAY-1998 (first entry)
XX
DE Thermotoga neapolitana endoglucanase DNA.
XX
KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW Biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase; ss.
XX
OS Thermotoga neapolitana (clone 56Gp1).
XX
PN W09744361-A1.
XX
PD 27-NOV-1997.
XX
PF 22-MAY-1997; 97WO-US08793.
XX
PR 22-MAY-1996; 96US-0651572.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Lam DE, Mathur EJ;
XX
PF WPI: 1998-018435/02.
DR P-PSDB; AAW34992.
XX
PT Endoglucanase(s), preferably form archaeal bacterium, AEP11 la -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose
XX
PS Claim 3; Fig 1H; 164pp; English.
XX
CC This DNA sequence from Thermotoga neapolitana (clone 56Gp1) encodes
CC an endoglucanase (see AAW34992) that is able to degrade
CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic
CC bonds in cellulose, and which shows homology to the thermostable
CC endoglucanase (see AAW34985) of archaeobacterium hydrothermal vent
CC isolate AEP11a. The DNA can be used in the recombinant production
CC of the endoglucanase and as a probe to identify similar sequences.
CC 24 Endoglucanase polynucleotides (see AAT94193-216) are claimed.
CC These can be incorporated into plasmid or virus-derived vectors for
CC use in a claimed method of producing enzymes in transformed host
CC cells. The claimed endoglucanases (see AAW34985-W35008) can be used
CC to degrade cellulose for the conversion of plant biomass into fuels
CC and chemicals, for use in detergents, textiles, animal feed, waste
CC treatment, and in the fruit juice and brewing industries for the
CC clarification and extraction of juices.
XX
SQ Sequence 2001 BP; 598 A; 394 C; 580 G; 429 T; 0 other;
Query Match 5.4%; Score 124.4; DB 19; Length 2001;
Best Local Similarity 53.0%; Pred. No. 1.2e-14;
Matches 424; Conservative 0; Mismatches 316; Indels 60; Gaps 5;
QY 376 CCGCGGTACAAACACCGCGCGCGCTTCCAGGCGCTTCACTACGGATCGGAGCGCG 435
Db 298 CCAGAGGGTACGAACCGCTCAGGACGGGTTTGAAGACTCGCACTACAGGTAGCGAAGCA 357
QY 436 GCGCGCACGCGCTTCGGGTGATGTGCTCTCACCACCACTGGAAGAAATTTGGGGGA 495
Db 358 AAAGAACTGGGCATCAATCGTCTTGTGNACCACTGGGACGATTCGGTGGGA 417
QY 496 ATGCATCAATACGAAGTGTACGGCGCTTCTTACACGACAACTTCTACACCGACCCC 555
Db 418 ATGAATCAATACGTGATGTGTTGGGGGCATCCATCAGCATG:..TCTACAGGAACGAG 477

QY 556 CGGACCCAGCAGGGGTACAAGAATTGGGTCAATCATCTACTGAACCGGTCAACAGCATTT 615
Db 478 AAGATCAAGAAGAATAACAAAAGTACGTGTCTTCCCTCATAAACAGGGTGAACACTTAC 537
QY 616 ACCGGGTGACGTACAAGAAGCATCCACAGATCTTTGGTTGGGAATTTGCCAATAGCGG 675
Db 538 ACGGTGTCTTCTTACAGGGAAGAGCCCACTCATATGGATGGAACTGCGGAACGAGGCC 597
QY 676 CGCTCGGTAGGAAGCGGCACATTACCAACCTCGGGCAGCTGCACCTCAGGCGGACCATTTCTC 735
Db 598 AGGTCTGAACGGAC-----AAGTCTGGTAAACACACTCGTT 633
QY 736 AACTGGGTGATCAAAATGTGCGCGTACGTCAAAAGCATAGACCCCTAACCATATGCTCTCG 795
Db 634 GAATGGGTAGAGAGATGAGTGCTTACATAAAGAGTCTGGATCCAAACCACTGGTGGC 593
QY 796 GTCGGGACGAAGGGTCTTACA-----TTGGGTCAACGAGGGAACGGCT 841
Db 694 GTGGGAGACGAGGATTTCTCAACAACTACGAAGGCTTTCAGACCTTACGGTGGAGAGCT 753
QY 842 G-GCCATACAACGACCCGTCGACGGGTGCACAAACATGCTCTTCTCCGTCTCAAGAAC 900
Db 754 GAGTGGGCTTACAACGGATGTCGCGTGTGACTGGAAGAGACTTCTGGAGATAGAGAC 813
QY 901 ATTGACTTTTGGCAGGTATCACTGTACCCGAATTACTGGGCCAGA-----AC 948
Db 814 GTGGATTTTGTACGTTCCTCTCTACCCCTCCCACTGGGTGTGAGGCCCTCAAACTAC 873
QY 949 GCGACTGGGGAACGAATGATCAAGGATCATATTTGCGAATGGCGCAGCATCGGCAAG 1008
Db 874 GCACAGTGGGGGCAAGTGGATAGAGATCACAATAAAGATCGCAAAAGAGGTTGCAAAA 933
QY 1009 CCGACCATTTCTGAAGAATTTCGGCTGGCAGACACCGG-----ACGGCATTCGGTC 1059
Db 934 CCCGTGTTCTGGAAGAGTACGGTATTTCCCAAAAGTCCCGCGTCAACAGGGTTGCCATT 993
QY 1060 TATCAGACGTGGACCCAGACTGTGCGTACGAACGAGCGGTGAAGCACTTCTGGATG 1119
Db 994 TACAAATTTGGAACGATCTGCTCTACAACTCGTGGAAACGGTGGCCATGTTCTGGATG 1053
QY 1120 CTCGCTGGGAATGTCAACGG 1139
Db 1054 CTCGAGGAATCGGTGAAG 1073
RESULT 13
AAQ13001
ID AAQ13001 standard; DNA; 2977 BP.
XX
AC AAQ13001;
XX
DT 14-OCT-1991 (first entry)
XX
DE Endol gene encoding endoglucanase.
XX
KW Cellulase activity; detergent; ds.
XX
OS Bacillus spp. NCIMB 40250.
XX
FH Key Location/Qualifiers
FT CDS 677..2779
FT sig_peptide /*tag= a
FT mat_peptide /*tag= b
FT 770..2776
PN W09110732-A.
XX
PD 25-JUL-1991.
XX
PF 18-JAN-1991; 91WO-DK00013.
XX
PR 19-JAN-1990; 90DK-0000164.

XX (NOVO) NOVO NORDISK A/S.
PA Jorgensen PL, Schulein M, Hansen C;
XX WPI: 1991-238020/32.
XX P-PSDB: AAR13227.
DR Enzyme exhibiting cellulase activity from *Bacillus* sp. - is an
XX endo-glucanase, esp. useful for harshness redn. of cotton-contg.
XX fabrics.
XX Claim 14; Page 81; 96pp; English.
XX
XX The DNA (SEQ ID NO: 1) was sequenced from plasmid pPL517 which was
CC prep'd. from genomic DNA isolated from *Bacillus* spp. PL236. The
CC sequence contains three potential initiation sites at positions
CC 677, 737 and 749, but only the ATG at 677 is preceded by a ribo-
CC some binding site. The DNA may be inserted into an expression
CC vector for prodn. of recombinant endoglucanase in *E. coli* cells.
CC The enzyme is useful as a cellulytic agent.
XX See also AAQ13002 and AAQ13003.
XX
XX Sequence 2977 BP; 775 A; 737 C; 821 G; 644 T; 0 other;
SQ

Query Match 5.0%; Score 114.8; DB 12; Length 2977;
Best Local Similarity 57.1%; Pred. No. 8.4e-13;
Matches 250; Conservative 0; Mismatches 182; Indels 6; Gaps 2;

OY 1382 TGCAGTACAGAACATGATTTCGGCGCGGTGATACACAGATCAACCGGGTCTCCAGT 1441
 ||||| ||| ||| ||| ||||| ||| |||
DB 148 TGCAGTACAGAGCGCGCGATACAAATGCAGCGCACACAGATCAAGCGCTTCAACA 207
 ||||| ||| ||| ||| ||||| ||| |||
OY 1442 TGGTGAATACCGGCTCGTCGTCGCTGGATTTCACACGGTCAGGTGCGGTACTGGTTCA 1501
 ||||| ||| ||| ||| ||||| ||| |||
DB 208 TCAAAACACGGTACTTCGGCTGTTGATTTAAGCACGCTCAAAATCGCTACTACTTCA 267
 ||||| ||| ||| ||| ||||| ||| |||
OY 1502 CCGCGGATGGTGGTGTGCTGCACACTGGTGTGACACTGTGACTGGCGGCGGATGGGGTGTG 1561
 ||||| ||| ||| ||| ||||| ||| |||
DB 268 CCAAGGATGGTTCGCGCGGTGAACGGCTGGATC--GACTGGCGCGAGCTCGCGCGCA 324
 ||||| ||| ||| ||| ||||| ||| |||
OY 1562 GGAATATCCGCGCTCGTTCGCTCGGTGAAACCGCGGACCGGCGGACACCTACC 1621
 ||||| ||| ||| ||| ||||| ||| |||
DB 325 GCACATTACAGATCTGTTTGGCAACCATACCTGGCACCAATTCGGATACGTGGAGC 384
 ||||| ||| ||| ||| ||||| ||| |||
OY 1622 TGCAGTTCTGTTCCACTCGTGGAACTGTTGGCGCTGGTGGGTGCGACGGTGAGATTCAAA 1681
 ||||| ||| ||| ||| ||||| ||| |||
DB 385 TGAGCTTCTGTCGAGCGCAGGCTCGATTGCGCGGCGGCGCAATCCCGTGAATCCAGC 444
 ||||| ||| ||| ||| ||||| ||| |||
OY 1682 ACCGGGTGAATAAGATGACTGTCGAACTTTGATGAGACCAANTGACTACTCGT---ATG 1738
 ||||| ||| ||| ||| ||||| ||| |||
DB 445 TGCGCATGTCACAGCGAGCTGGTCGAACTTTACGAGCGGACGACTACTGTCGTATG 504
 ||||| ||| ||| ||| ||||| ||| |||
OY 1739 GGACGAACACCGCTTCCAGGATTTGGACGAAGGTGACGGTGTATGTCAATGGCGGGTGG 1798
 ||||| ||| ||| ||| ||||| ||| |||
DB 505 GGACGAAGACGGCCTTTGCTGACTGGGATCGGTGCGTATTGTACAGAACGGCCCAATAG 564
 ||||| ||| ||| ||| ||||| ||| |||
OY 1799 TGTGGGGAGTGAACCGT 1816
 ||||| ||| |||
DB 565 TGTGGGGAAGTGTCCCAT 582
 ||||| ||| |||

RESULT 14
AAV07163
ID AAV07163 standard; DNA; 1621 BP.
XX
XX AAV07163;
AC
XX
XX 11-SEP-1998 (first entry)
XX
XX Thermomonospora fusca cellulase E2 gene.
DE
XX cellulase; E2; recombinant plant; cellulose-degrading enzyme;
KW

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2002, 08:52:51 ; Search time 2135 Seconds
(without alignments)
17363.686 Million cell updates/sec

Title: US-09-917-378-2
Perfect score: 2289
Sequence: 1 atgddgtactgctgcgcc.....tcactgcacgctagtgtga 2289

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estov.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_fod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81.2	3.5	974	17	BH157536	ENTRL88TR
2	80.8	3.5	974	17	BH157536	ENTRL88TR
3	79.4	3.5	335	17	A2415847	A2415847
4	79.2	3.5	463	13	B1781048	EBma03_SO
5	79.2	3.5	678	14	BQ760233	EBp107_SO
6	78.6	3.4	591	12	BG240228	OVI19_DI

7	78.6	3.4	655	13	BJ248065	BJ248065
8	76.4	3.3	1315	10	AV752548	AV752548
9	76.2	3.3	788	17	CNS04AAV	AL281632 Tetraodon
10	76	3.3	236	9	AA720369	AA720369 33562 Lam
11	75.6	3.3	586	17	A2871772	AZ871772 2M0184M17
12	75.2	3.3	939	17	AG043613	AG043613 Pan trogl
13	74.4	3.3	829	17	CNS035GN	AL228704 Tetraodon
14	73.6	3.2	818	17	CNS044N4	AL274297 Tetraodon
15	72.4	3.2	616	17	A2432179	A2432179 IM0217A05
16	72	3.1	988	17	CNS022KP	AL178306 Tetraodon
17	71.2	3.1	135	9	AL820149	AL820149 AL820149
18	71	3.1	955	17	CNS02RNO	AL210813 Tetraodon
19	70.8	3.1	865	12	BF315505	BF315505 601899637
20	70.6	3.1	162	12	BE918956	BE918956 FM1_2_C07
21	70.6	3.1	925	17	CNS00CAZ	AL058951 Drosoph11
22	70.2	3.1	464	12	BE799902	BE799902 60158045
23	70.2	3.1	622	13	B1994975	B1994975 1031018D0
24	69.6	3.0	562	17	A2855343	A2855343 2M0159A22
25	69.6	3.0	610	17	A2369884	A2369884 IM0120111
26	69.4	3.0	559	17	BH306809	BH306809 CH230-100
27	69.4	3.0	791	17	CNS028QT	AL186302 Tetraodon
28	69.2	3.0	724	17	A2197686	A2197686 SP_1035_B
29	69.2	3.0	2003	12	BG330370	BG330370 602430456
30	68.8	3.0	464	17	A2928926	A2928926 479.d1f19
31	68.2	3.0	438	17	A2858646	A2858646 2M0163J20
32	67.8	3.0	843	17	CNS00CS1	AL059666 Drosoph11
33	67.8	3.0	875	17	AG043475	AG043475 Pan trogl
34	67.6	3.0	826	17	BH316049	BH316049 CH230-40E
35	67.4	2.9	927	17	AG127811	AG127811 Pan trogl
36	67.4	2.9	955	17	CNS02YA4	AL219397 Tetraodon
37	67.2	2.9	527	17	A2927141	A2927141 RPCL-23-1
38	67.2	2.9	565	17	A2386138	A2386138 IM0144P14
39	67.2	2.9	959	17	CNS022KT	AL221078 Tetraodon
40	67	2.9	828	17	CNS04XH7	AL311668 Tetraodon
41	66.8	2.9	1160	17	AG043473	AG043473 Pan trogl
42	66.6	2.9	402	14	BQ568798	BQ568798 g1114h07.
43	66.6	2.9	600	12	BG802543	BG802543 0182-32 M
44	66.6	2.9	804	13	B1946620	B1946620 HVSMF1000
45	66.6	2.9	976	17	A2668903	A2668903 ENTIC14TR

ALIGNMENTS

RESULT 1
BH157536
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BH157536
ENTRL88TR Entamoeba histolytica Sheared DNA
genomic, DNA sequence.
BH157536
BH157536.1 GI:15730974
GSS
Entamoeba histolytica.
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 974)
Loftus B., Wang Z., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 68
High quality sequence stop: 206.

Class: shotgun

High quality sequence start: 68

High quality sequence stop: 206.

FEATURES

source

Location/Qualifiers

1..974

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/notes="Vector: pHOSt; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (-2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

BASE COUNT

238 a 212 c 291 g 233 t

ORIGIN

Query Match

Best Local Similarity

Matches 161;

Conservative

Score 81.2;

DB 17;

Length 974;

Pred. No. 2.9e-06;

Mismatches 133;

Indels

0;

Caps

0;

Qy

1819

GCCACAGCCCCAGCCCACACAGCCGCGCAGCCCAAGCCGTCTCCCGAGCCCGAGCCGC 1878

|||||

Db

657

GCGACATACACCC 716

Qy

1879

ACCCAGACCCAGCTGCTCCGCCATCCCGTCGCCGAGCCGACGCGCCAGCCCTACGCGC 1938

|||||

Db

717

CC 776

Qy

1939

TCCCCGTGCGCGAGCCGCTCCCGTCGCCGAGTGTCGTCTGGGTGTGGGTGCGCG 1998

|||||

Db

777

CCCCCCCCCCCCCCCCCCCCCCCCACAGAGGGGGGGGGGGGGGGGGGGGGGGGGGG 836

Qy

1999

CGGAGGTATGTGTTAATAGTAGTTGGGTCTTGGGTTTACGGCGACGGTGACGTGACG 2058

|||||

Db

837

GG 896

Qy

2059

AATACCGGAGCCGGGACCAGCGGTGGACGGTGCGGTGCTGTTTGGTGGG 2112

|||||

Db

897

GG 950

RESULT 2

BH157536/c

LOCUS

DEFINITION

ENTRU88TR Entamoeba histolytica Sheared DNA linear GSS 24-SEP-2001 genomic, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1

(bases 1 to 974)

Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse

Class: shotgun

High quality sequence start: 68

High quality sequence stop: 206.

FEATURES

source

Location/Qualifiers

1..974

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/notes="Vector: pHOSt; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (-2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

BASE COUNT

238 a 212 c 291 g 233 t

ORIGIN

Query Match

Best Local Similarity

Matches 163;

Conservative

Score 80.8;

DB 17;

Length 974;

Pred. No. 3.5e-06;

Mismatches 137;

Indels

0;

Caps

0;

Qy

1813

CCGTCCGGCACACAGCCGCGCAGCCGCGCAGCCGCGCAGCCGCGTCCCGAGCCGC 1872

|||||

Db

972

CC 913

Qy

1873

AGCCGACCCCAAGCCGAGCTCTCTCCCATCCCGTCGCCGAGCCGACGCGCCAGCCCT 1932

|||||

Db

912

CC 853

Qy

1933

ACGCCGTCCCGCTCGCCGAGCCGCTCCCGTCGCCGAGTGTCGTCTGGGTGTGGGG 1992

|||||

Db

852

CC 793

Qy

1993

TGCCGGCGGAGGTATGTGTTAATAGTAGTTGGGTCTTGGGTTTACGGCGACGTTGACG 2052

|||||

Db

792

GGCGGG 733

Qy

2053

GTGACGAATACCGGGACCGCGGACGACGCGGTGGGTGGGTGCTCTTTTGGTGGG 2112

|||||

Db

732

GG 673

RESULT 3

AZ415847/c

LOCUS

DEFINITION

1M0190C20R Mouse 10kb plasmid UUGCIW library Mus musculus genomic clone UUGCIW0190C20 R, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1

(bases 1 to 335)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah

Query Match	3.2%	Score 73.6;	DB 17;	Length 818;
Best Local Similarity	68.5%;	pred. No. 9.5e-05;		
Matches 100;	Conservative 1;	Mismatches 45;	Indels 0;	Gaps 0;
Qy	1831	AGCCCCACACCCAGCCGACGCCAACCCGCTGCCGAGCCGAGCCGACCCGACCCGACCCGCCC 1890		
Db	817	ACCTGACCAACCAACCTTACCCAAACCCGACCCGACCCGCGGCGCCGACCCGACCCGACCCG 758		
Qy	1891	AGCTCCTCCGACATCCCGTCCCGAGCCGCCAGCCGACCCGACCCGCTACGCGTCCGCTGCGCCG 1950		
Db	757	ACCTGACCCCTTAACCCCTTGACCCCAACCTGTACCCCAACCCGACCCGCGGCGCTGACCCCG 698		
Qy	1951	AGCCCGTCCGCGTCCGCGAGTGTGTC 1976		
Db	697	ACCCGACCCGACCCGACCCCTGAC 672		
RESULT 15				
AZ432179/c				
LOCUS	AZ432179	616 bp	DNA	linear
DEFINITION	1M0217A05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0217A05 R, DNA sequence.			GSS 03-OCT-2000
ACCESSION	AZ432179			
VERSION	AZ432179.1	GI:10556192		
KEYWORDS	GSS.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 616) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly M., Rose,M., Rose,R., Slokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.			
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., StG, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0217 row: A column: 05 Seq primer: CACACAGGAACACGATGACC Class: plasmid ends High quality sequence stop: 616. Location/Qualifiers 1..616 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="UUGC1M0217A05" /clone_lib="Mouse 10kb plasmid UUGC1M library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, T1-resistant, P-" /note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114 gb AF129072.1), a copy-number			
FEATURES				
source				

Db 2078 CCAGGCGGTTCGGAGTCCGAGCCGACGCTACTCCGACGCGACAGCCGCGACG 2137
QY 1912 CCAGGCGCGAGCCAGCCCTAGCGGTCCCGTCCGCGAGCCGCGTCCGCGAGT 1971
Db 2138 CCAACGCTGACCCCTACTGCTAGCCGACCCGCGAGCCGCGAGCCGCGTCCCGA-- 2195
QY 1972 GTCTGCTGCTGCGGTTCGGGTCGCGGCGACGCTATGCTGTAATAGTATTGGGGTCT 2031
Db 2196 -CGGACGCTCCGAGCCGCTGCGCGGAGTTACAGGTCAACAGGATTGGGCAAT 2254
QY 2032 GGGTTTACGGCGAGCGGTGAGCGGTACAGAAATACCGGAGCCGCGGAGCGCGGTGAGC 2091
Db 2255 GCGTTACCGTAACGCGTGGCGGTGACAAATTCGCGATCCGTCGCGCAAGACATGGACG 2314
QY 2092 GTGCGGTGGTCTGTTGGTGGAAATCAGCGGTACGAACTACTGGAACTGCGTTGACC 2151
Db 2315 GTGAGTTGGAACTTCGGCGGAATCAGACGATTACCAATTCGTTGGAATGCGCGGTACG 2374
QY 2152 CAATCAGGTGCATCGGTGACGGCGAGCACTGAGCTTACAAACGCTGATCCAAACCGGT 2211
Db 2375 CAGAAGGTCAGTCGTTACGCTGCGGATATGAGTTATAACAACGTTGATTCAGCCTGGT 2434
QY 2212 CAGTCGACCACTTCGGATTCAACGGAATTTACTCAGGAACAACACCGGCACCTACACTC 2271
Db 2435 CAGAACAACCACTTCGGATTTCAGGCGAGCTATACCGGAAGCAACGCGGCGACGACATC 2494
QY 2272 ACCTGACGCGCTAGTTGA 2289
Db 2495 GCCTGCGCAACGATTAA 2512

RESULT 2

US-09-136-574A-1

; Sequence 1, Application US/09136574A

; Patent No. 6284366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Berquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; Name: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

;

; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-136-574A-1

Query Match 5.8%; Score 133.6; DB 4; Length 11707;
Best Local Similarity 52.4%; Pred. No. 4.2e-20;
Matches 395; Conservative 0; Mismatches 344; Indels 15; Gaps 4;

QY 1220 CAATCAGTACCGGCGACATCGCCTCCGCGGTTCGCGGAGCTCGATCCATCTCTGTCGCGGT 1279
Db 3883 CAACACACCTGCACCGACATCAAGCCCGGACAGCCGCAACAGTACACAGCAACGCGGA 3942
QY 1280 CTCGTCGCGGTCTCCGTCGCGGTCTCCGTCGCGGTCTCCGTCGCGGTCTCTTCTGTCGCGGA 1339
Db 3943 CTCACGCGCGCGCGACAGTACGCTTACTGTGACTCCGACACCGACACACACCGA 4002
QY 1340 GCCGTCCTCGTCGTCGTCGCGGTCTCGG---GTGGGGTGAAGGTCAGTACAAGAACA 1396
Db 4003 CGCGCACCGACAGGACACCTTGGCACGGAGTGGTTTGAAGTACTATACAAGAACA 4062
QY 1397 ATGATTCCGCGCGGTGATACACAGATCAACCGGGTCTCCAGTTCGTTGAATACGGGGT 1456
Db 4063 ATGAGCAAGTCGCGAGCACAAAGTTCTATAAGGCGGTGGTTAAGATAGTGAATGGAGGGA 4122
QY 1457 CGTCGTCGTTGATTTGTCGACGCTGACGCTGCGGTACTGTTCCACCGGGATCGTGGGT 1516
Db 4123 GCAGCAGTGTGATCTTAGCAGGGTTAAGATAAGATACTGTTACACAGTGGATGCTG 4179
QY 1517 CGTCGACACTGTTGTACAACTGTGACTGGCGGGGATGGGTGGGAATATCCGCCCTT 1576
Db 4180 ACAAGCCACAGAGTGGCGGTATGTACTGGGCACAGATAGGGGCAAGCAATGTGACATTCA 4239
QY 1577 CGTTGGCTCGGTGAACCGCGGACGCGGACGCGGACACTTACCTCGAGTTGTGCTTCA 1636
Db 4240 ATTTGTGAAGCTGAGCAGCGGAGTGAAGTGGAGGATATTACTTGGAGTAGGATTTA 4299
QY 1637 CTGCTGGAAC-----GTTCGCGCGCTGGTGGTTCGACGGGTGAGATTCAAAACCGGGTGA 1690
Db 4300 CGAGTGGAGCTGGGACGTTCCAGCCTTGGTAAGGACACAGGGGATATACAGGTAAGTTTA 4359
QY 1691 ATAGAGTACTGTCGAACCTTTGATGAGACCAATGACTACTGTTGGGACGAAACCG 1750
Db 4360 ACAAGATGACTGGAGCAATTACAATCAGGCACGACTGCTCATGTTGCAGAGCATGA 4419
QY 1751 CCTTCCAGGATTGGA---CGAAGTGAAGTGTATGTCAATGGCGGCTGTTGGGGGA 1807
Db 4420 CGAATTATGAGAGATGCGAAGTAAAGCTGTATGTAGATGGTGTCTGGTATGGGGG 4479
QY 1808 CTGAACGTCGCGACACAGCCCGACAGCCCGACAGCCCGACAGCCCGACAGCCCGACAGCCCGA 1867
Db 4480 AGGAGCGCGCGGACGACACCTTGCACCGACAAGCAGCAGCAACCAACCAACCAACCAACCA 4539
QY 1868 GCCCGAGCGCGACCCCGACAGCCCGACAGCTCTCTCCCATCCCGCTCCCGAGCGCCCGACCA 1927
Db 4540 CAGCAACAGCAACACCGCGCGACAGCAACGCGCAAGCTCTACACCGACACACGACAGCA 4599
QY 1928 GCCTACGCGCTCCCGGTTCGCGGAGCGCGGTCCGCT 1961
Db 4600 CACCAACCCCAATACCAACACACCCCAACGCGCTCC 4633

RESULT 3

US-09-136-574A-2

; Sequence 2, Application US/09136574A

; Patent No. 6284366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6416 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2
Query Match 5.6%; Score 128.4; DB 4; Length 6416;
Best Local Similarity 52.6%; Pred. No. 4.9e-19;
Matches 379; Conservative 0; Mismatches 326; Indels 15; Gaps 4;
QY 1243 CCGCCGTCGCGGAGTCGAGTCCTCTCGTCGCGGTCCTCGTCGCGGTCCTCGTCGCGG 1302
Db 2533 CCGGGTCGGGCAACATCCACACCTGCACCAACATCAACACCAACACCGGACCCCA 2592
QY 1303 TCCTCGGTCGCGGTCCTCGTCGCGGAGCCGTCCTCGTCGTCGTCGTCGTCGCGG 1362
Db 2593 ACACCAACAGTGACACGACGCGGACGCGGACCTCTACACCGACACGCGGTCACCT 2652
QY 1363 GTCTCGG---GTGGGGTGAAGTGCACTACAAGAACAAATGTCGCGCGCGGTGATAAC 1419
Db 2653 GGTACGGGAAGTGTGTGAAGTACTGTACAAGAACAAATGAGACAGTGGGACACAGGT 2712
QY 1420 CAGATCAACCGGTCCTCCAGTTGGTGAATACCGGGTCGTCGTCGTCGTCGTCGTCG 1479
Db 2713 TCTATAAGCCGCTGGTTAAGATAGTGAATGGAGGACGACAGTGTGTGATCTTAGCAGG 2772
QY 1480 GTACCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1539
Db 2773 GTTAAGATAAGTACTGTTGACACAGTGGATGGTG---ACAAGCCACAGAGTGGGATGT 2829
QY 1540 GACTGGCGCGGATGGGTCGTGGGAATATCCCGCGCTCTGGGTCGTCGTCGTCGTCGTCG 1599
Db 2830 GACTGGGCACATAGGGGCAAGCAATGTGACATTTCAATTTTGTGAAGCTTACCGCGGA 2889

QY 1600 ACGCCGACGGCGGACACCTACCTGCAGTTGCTGTTCACTGCTGGAAC-----GTTGGCC 1653
Db 2890 GTGAGTCGAGCGGATATTACCTGGAGCTAGGATTTACGAGCTGGGAGCTTGCAG 2949
QY 1654 GCTGCTGGTCGACGGGTGAGATTCAAAACCGGGTGAATAAGAGTGACTGTCGAACTTT 1713
Db 2950 CTTGTAAAGACACAGGGGATATACAGGTAAGGTTTAAAGAAGATGACTGGAGCAATTAC 3009
QY 1714 GATCAGACCAATGACTACTCTGATGGGACGAACACCGCTTCCAGGATTGA---CGAAG 1770
Db 3010 AATCAGGACGACGACTGGTCATGGTTGCAGAGCATGACGAATTTATGGAGAGAATCCGAAG 3069
QY 1771 GTGACGGTGTATGTCAATGGCCGCTGCTGTTGGGGGACTGAACCGTCCGGACACCGCC 1830
Db 3070 GTGACGCTGTATGTAGATGGTGTCTGCTATGGGGCAGGAGCCGGGAGGAGGACACCT 3129
QY 1831 AGCCCCACACCCAGCCCGCCAGCCCAACCCGTCCTCCCGAGCCCGAGCCCGACCCAGCC 1890
Db 3130 GCACCGACAAGCACAGCAACACCAACGCCAACTCCGACAGCAACCCCAACCTACACCT 3189
QY 1891 AGCTCTCCCGCATCCCGTCCCGCGAGCCCGCCAGCCCGCTAGCCGCTCCCGCTCCCG 1950
Db 3190 ACACCGACCCGACACCGACAGTGTGTCACAGCCGACACCGCGGACCGGATCACCG 3249
RESULT 4
US-09-134-078-57
Sequence 57, Application US/09134078
Patent No. 638844
GENERAL INFORMATION:
APPLICANT: Bylina, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 1992 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...1989
US-09-134-078-57

Query Match 5.5%: Score 126.4; DB 4; Length 1992; Best Local Similarity 49.3%; Pred. No. 9.9e-19; Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;									
QY	121	TTCTGTCACCGCATCCGGCGGTCACTTCGTCTGAACCGCCTTCCTATCGTTACGGGGGA	180						
DB	49	TTCTGTAAGTGGAAACCGGAAATTCGCTCTGAACGAAAGAAATTCAGATTCATTGGA	108						
QY	181	ACGAACAACATATACCTCAGCTATCAGTCGCACCGCAGCTCGATGACGTGTTCGCCAAG	240						
DB	109	AGCAACAACATACATACATGCACTACAAAGCAGCAAGCAATGATAGACAGTGTTCGAGAGT	168						
QY	241	GCTCAACGCGATGAATCTTTCTGTCACTCCGACCTGGGTTTCATCGACATCGGCTCTCTT	300						
DB	169	GCCAGACACATGGGTATAAAGGTCTCAGAACTCTGGGTTCCTTCGACGGGAGAGTTAC	228						
QY	301	GACGGCTCCGTGCCCAATCGATGGCAACAAGAACCGCTTCTACTTTCAGTACTGGGAC	360						
DB	229	TGCAGAGACAAGAACACTACATGCTCTGAGCCCGGTGTTCGGGGTGCCAGAGGA	288						
QY	361	CCGTGACCGCGCTCCGGGCTACAACGCGCGCCGCTTGAAGGCTTGAAGTAC	420						
DB	289	ATATCGAACGCC-----CAGAGCGGTTTCGAAAGACTCGACTAC	327						
QY	421	GCGATCCGAGCGCGCGCGCAGCGCTTCGGGTGATGTGCTCTCAACCAAGACTGG	480						
DB	328	ACAGTTGCGAAAGCGAAAGAACTCGGTATAAACTTCTCATTTGTTGTGAACAACCTGG	387						
QY	481	AAGAAATTCGGGAATGGATCAATACGACAGTGTACCGCTTCCTTACCAGCACAAC	540						
DB	388	GACGACTTCGGTGAATGAACAGTACGTAGGTGGTTGGGAGAACCCATCAGAGCAT	447						
QY	541	TTCTACACCGACCCCGACCCAGCGGTACAAGAAATTTGGGTCAATCATCTACTGAAC	600						
DB	448	TTCTACAGAGATGAGAAGATCAAGAAGAGTACAAAAAGTAGCTCTTCTCTGTAAC	507						
QY	601	CGGTCAACAGCATTACCGGGTGACGTACAAGAACGATCAAGCATCTTTGCTTGGGAA	660						
DB	508	CATGTCTAATACCTACACGGGAGTTCTTACAGGAAGAGGCCACCATCATGCGCTGGGAG	567						
QY	661	CTTGCCAAATGAGCGCGCTCGTAGGAAGCGGCACATTACCAACCTCGGCACGCTGCAC	720						
DB	568	CTTGCAACCAACCGCGCTGTAGACGCGACAAATCGGGGACACGCTCTGTGAGTGGGTG	627						
QY	721	CAGCGACCATTTGTAACCTGGGTGATCAAAATGTGCGGTACGTCAAAAGCATAGACCT	780						
DB	628	AAGGAGA-----TGAGTCTCTACATAAAGAGTCTGCATCCCAACCACTCGTGGCTGTG	681						
QY	781	AACCATATGCTCTCGGTGCGGACGAGGTTCTACATTTGGGTCAACGCGAGGAGCGGC	840						
DB	682	GGGACCAAGGATTTCTTCAGCAACTACGAAGGATTTCAAACTTTACGGTGGAGAGCCGAG	741						
QY	841	TGGCCATACAAACCGCTCCGACGGGTCCGAAACAATGCTCTCTCCGTGTCAAGAAC	900						
DB	742	TGGGCTTACAAACCGGTGTC---CGGTGTGACTGGAAGAGTCTCTTCGATAGAGCG	798						
QY	901	ATTGACTTTGGCAGTATACCTGTATCCCGAATTTACTGGGG-----CCAGAAC	948						
DB	799	GTGACTTCGCCACGTTTCCACCTCTATCCGTCCCACTGGGGTGTCACTCCAGAGAACTAT	858						
QY	949	GCGGACTGGGGAACGCAATGATCAAGGATCATATTCGGAAATCCGCGAGCGGATCGGCAAG	1008						
DB	859	GCCAGTGGGGAGCGAAGTGGATAGAAGACCAATAAAGATCGCAAAAGAGATCGGAAAA	918						
QY	1009	CCGACCATTTCTGAAGAATTTCCG-----CTGGCAGACACCGGACCGGATTCGCTC	1059						
DB	919	CCCGTTGTTCTGGAAGAATATGNAATTCCAAGAGTCCGCCAGTTACACAACGGCCATC	978						
QY	1060	TATCAGACGTGGACCAAGCTGTGCGGTACGAACGGGTGAACAGCGTGGAACTTCTGGATG	1119						
DB	979	TACAGACTCTGGGAACGATCTGGTCTACGATCTCGGTGGAGATGGAGCATGTTCTGGATG	1038						
QY	1120	CTCGCTGGGAATCTCAACGG	1139						

DB	1039	CTCGCGGAATCGGGAAGG	1058						
RESULT 5									
US-09-134-078-11									
Sequence 11, Application US/09134078									
Patent No. 636844									
GENERAL INFORMATION:									
APPLICANT: Bylina, Edward J.									
TITLE OF INVENTION: GLYCOSIDASE ENZYMES									
NUMBER OF SEQUENCES: 72									
CORRESPONDENCE ADDRESS:									
ADDRESSEE: Gray Cary Ware & Freidenrich LLP									
STREET: 4365 Executive Drive, Suite 1600									
CITY: San Diego									
STATE: CA									
COUNTRY: USA									
ZIP: 92121									
COMPUTER READABLE FORM:									
MEDIUM TYPE: Diskette									
COMPUTER: IBM Compatible									
OPERATING SYSTEM: Windows95									
SOFTWARE: FastSeq for Windows Version 2.0									
CURRENT APPLICATION DATA:									
APPLICATION NUMBER: US/09/134.078									
FILING DATE: 13-AUG-1998									
CLASSIFICATION: 435									
PRIOR APPLICATION DATA:									
APPLICATION NUMBER: 08/949,026									
FILING DATE: 10-OCT-1997									
APPLICATION NUMBER: 60/056,916									
FILING DATE: 06-DEC-1996									
ATTORNEY/AGENT INFORMATION:									
NAME: Halle, Lisa A.									
REGISTRATION NUMBER: 38,347									
REFERENCE/DOCKET NUMBER: 09010/024002									
TELECOMMUNICATION INFORMATION:									
TELEPHONE: 858/677-1456									
TELEFAX: 858/677-1465									
INFORMATION FOR SEQ ID NO: 11:									
SEQUENCE CHARACTERISTICS:									
LENGTH: 2043 base pairs									
TYPE: nucleic acid									
STRANDEDNESS: double									
TOPOLOGY: linear									
MOLECULE TYPE: Genomic DNA									
FEATURE:									
NAME/KEY: Coding Sequence									
LOCATION: 1...2040									
US-09-134-078-11									
Query Match 5.5%: Score 126.4; DB 4; Length 2043; Best Local Similarity 49.3%; Pred. No. 9.9e-19; Matches 513; Conservative 0; Mismatches 476; Indels 51; Gaps 5;									
QY	121	TTCTCTACCCCATCCGGCGGTCACTTCGTCTGAACCGCCTTCCTCTATCGTTACGGGGGA	180						
DB	100	TTCTGTAAGTGGAAACCGGAAATTCGCTCTGAACGAAAGAAATTCAGATTCATTGGA	159						
QY	181	ACGAACAACATTTACCTCAGCTATCAGTCGCACCGCAGCTCGATGACGTGTTCGCCAAG	240						
DB	160	AGCAACAACATCTACATGCACTACAAGCAACGCAATGATAGACAGTGTTCGAGAGT	219						
QY	241	GCTCAACGCGATGAATCTTTCTGTGTCATCCGACCTGGGTTTCATCGACATCGGCTCTCTT	300						
DB	220	GCCAGAGACATGGGTATAAAGGTCTCAGAACTCTGGGTTTCCTTCGACGGGAGAGTTAC	279						
QY	301	GACGGCTCCGTGCCCAATCGATGGCAACAAGAACCGCTTCTACTTTCAGTACTGGGAC	360						
DB	280	TGACAGACAGAAGAACACCTACATGCATCTCGAGCCCGGTGTTTTCCGGGGTCCCAAGGA	339						
QY	361	CCGTGACCGCGCTCCGGGCTACAACGCGCGCCGCTTGAAGGCTTGAAGTAC	420						

Db 340 ATATCCAAACGCC-----CAGAGCGGTTTCCAAAGACTCGACTAC 378
QY 421 CGGATCGGCGGCGCGCGCGCGCTTCGGGTGATTCGTCCTCACCACGACTCG 480
Db 379 ACAGTTGCCAAGCGAAGAACTCGGTATAAACTTGTCTATTGTTCTTGAACTCG 438
QY 481 AAAGAATTTGGGGGAATGGATCAATACGACAAAGTGTACGGCTTCTCTTACCACGACAA 540
Db 439 GACGACTTCGGTGGAAATGAACCACTGAGTGGTGGTGGAGAACCCATCAGCAGAT 498
QY 541 TTCTACACGACCCCGGCGGCGGCGGTACAGAAATTTGGTCAATCATCTACTGAAC 600
Db 499 TTCTACAGAGATGAGAGATCAAGAGAGATCAAAAGTAGCTCTCTTTCTCGTAAAC 558
QY 601 CGGGTCAACAGCATTTACCGCGGTGACGTACAAGAACGATCCAAAGCATTTTCTGGGAA 660
Db 559 CATGTCAATACCTACACGGGAGTTCTTTACAGGAAGAGCCCAACCATCATGGCTGGAG 618
QY 661 CTTGCCAATGACCGCGGCTCGGTAGGAAGCGGCACATTTACCAACCTCGGGCAGTGCAT 720
Db 619 CTTGCAACGAACCGCGCTGTGAGCGGACAAATCGGGGAACACGCTCTTGAAGTGGTG 678
QY 721 CAGGCGACCATTTCAACTGGGTGCGATCAATCTCGCGCTACGTCAAAAGCATAGACCT 780
Db 679 AAGGAGA-----TGACTCTCTACATAAAGAGTCTGGATCCCAACCACTCGTGGCTGTG 732
QY 781 AACCATATGCTCGGTGCGGCGGCGGAGGTTCTACATTTGGGTCAACGCGGAGCGGC 840
Db 733 GGGGACGAGGATTTCTTACGAACTACGAAAGGATTCAAACCTTACGCTGGAAGCGGAG 792
QY 841 TGGCCATACAGCACCGCTCGGCGGCGGTGCGACAAATGCTCTTCTCGGTGTCAAGAAC 900
Db 793 TGGGCTTACAACGGGTGGTTC---CGGTGTGACTGGAAGAAGCTCTTTTCGATAGAGCG 849
QY 901 ATTGACTTTGGGACGATATCACTGTACCCGTAATTTACTGGG-----CCAGAC 948
Db 850 GTGGACTTCGGGACGTTCCACCTCTATCTCGTCCGCTGCGGTGTCAGTCCAGAGAACTAT 909
QY 949 GCGGACTGGGGAACCAATGGATCAAGGATCATATTTGCGAATCCCGCAGCGATCGGCAAG 1008
Db 910 GCCCAGTGGGAGCGAAGTGGATAGAGACCAACATCAAAAGATCGCAAAAGATCGGAAA 969
QY 1009 CCGACCATCTTCAGAAATTCGG-----CTGGCAGACCGGACCGCGATTCGGTC 1059
Db 970 CCGGTGTGTTCTGGAAGAATATGGAATTCGGAAGAGTGGCCAGTTAACAGAACGCGCATC 1029
QY 1060 TATCAGACTGGACCCAGACTCTGCTACGAACGGTGAAGCAGGCTGGAAGTCTGGATG 1119
Db 1030 TACAGACTCTGGAACGATCTGCTCTACGATCTCGGTGGAGATGGAGCGATGTTCTGGATG 1089
QY 1120 CTCGCTGGGAATGTCAACGG 1139
Db 1090 CTCGCGGGAATCGGGAGG 1109

RESULT 6

US-08-525-697-1

: Sequence 1, Application US/08525697

: Patent No. 5795764

: GENERAL INFORMATION:

: APPLICANT: Christgau, Stephan

: APPLICANT: Andersen, Lene N

: APPLICANT: Kauppinen, Sakari

: APPLICANT: Heldt-Hansen, Hans P

: APPLICANT: Dalboe, Henrik

: TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY

: NUMBER OF SEQUENCES: 15

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: No. 57957640 No. 5795764disk of No. 5795764th America, Inc.

: STREET: 405 Lexington Avenue, 64th Floor

: CITY: New York

: STATE: New York

: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/525,697
: FILING DATE: 21-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Harrington, James J.
: REGISTRATION NUMBER: 38,711
: REFERENCE/DOCKET NUMBER: 4004,204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK 0486/93
: FILING DATE: 30-APR-1993
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1302 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 61..1192
: US-08-525-697-1

Query Match 5.2%; Score 118.6; DB 1; Length 1302;

Best Local Similarity 49.6%; Pred. No. 4,4e-17;

Matches 436; Conservative 0; Mismatches 389; Indels 54; Gaps 3;

QY 351 GTACTGGGACCGGCTCGGCGGTACAAAGCGGCGGCGGCGGCGGCTTGCAGG 410
Db 357 GTACAGCTGCAGCGGACGCGACCTCGACCATCAACAGGCGGCGGCGGCTCCAGCG 416
QY 411 CTTTGAATACGGGATCGGAGCGGCGGCGGCGGCGGCTTGGGGTGAATGTCGCTCAC 470
Db 417 CCTGACTAGTGGTGCCTCGGCGGAGAGTACGGCGTCAAGCTGATCATCAACTCGT 476
QY 471 CAACGACTGGAAGAAATTTGGGGAATGGATCAATACGACAACTGGTACGGCCTTCTTA 530
Db 477 CAACGAGTGCAGCGGCTACGCGGCGATGAGGCTACGTCACGGCGCTACGGCGCGCGCG 536
QY 531 CCACGACAACTTCTACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTGGGTCAATCA 590
Db 537 CCAGACGGACTTCTACACCAACACCGCGCATCCAGGCGGCGCTACAGAACTATACAGCG 596
QY 591 TCTACTGAACCGGCTCAACAGCATTTACCGGCGGTGAGCTACAAGACGATCCCAACGATCT 650
Db 597 GGTGCTGTCGCGGTACAGCAGCTCCCGCG-----CCATCTT 632
QY 651 TGTGGGAACTTGCCAATGAGCGCGCTGCGTGAAGCGGCGGCGGCGGCGGCGGCGG 710
Db 633 CGCGTGGGAGCTGGCGCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 679
QY 711 CAGGTGCACTCAGGCGGACCATTTGTCAACTGGGTGATCAATGTCGGCGGTACGTCGCTCA 770
Db 680 -----CGGTCTGTACAACTGGATCTCGGACAGCTCCCAAGTATATCAAGTC 725
QY 771 CATAGACCTTAACCATATGCTGCTCGGTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 830
Db 726 GCTGAGCTCCAAGCACTGCTGCGGATTCAGGCTTTCGCGTCTCGGCTCGGCTCGGCTCG 785
QY 831 GGAAGCGGCTGGCGCATACAACGCGGCTCCGACGCGGCTCGGCGGCGGCGGCGGCGGCGG 890
Db 786 CGACGCGGCGCTACCCCTACACCTAC-----GGCGAGGAGTTGAACCTTCCACCAAGAACCTGGG 842

Qy 891 TGTCAAGAACATTTGACCTTTGGCAGCTATCACCTGTACCCGAATTAATCTAGTGGGCGAGAACGC 950
Db 843 CATCTCGACCATCGACTTCGGCTACTCTGCATCTGTACCCCGATAGCTGGGCGACCTCTCTA 902
Qy 951 GGACTGGGGAACCAATGATCAAGGATCATATTCCGAATCCCGAGGATCGGCAAGCC 1010
Db 903 CGACTGGGGCAACGGCTGGATGATCAGCGGCCCGCCCTGCAAGCGCGTGGGCAAGCC 962
Qy 1011 CACCATTTCTCAAGAATTCGGCTGCGACAGCACCGGACCGGATTCGGTCTATCAAGACCTG 1070
Db 963 GTGCTGCTGAAGAGTACGAGTACCTTCCAAACCACTGTGCGCTCGAGAGCCCTTGCA 1022
Qy 1071 GACCCAGACTGTGGCTACGAACGGTGAAGCAGGCTGGAACCTTCTGGATGCTCGCTGGGAA 1130
Db 1023 GCAGACGGCGGCAACGCGAGCGGCGATCTCCGGCGATTGTGCTGCGAGTATGGCAACC 1082
Qy 1131 TGTCAACGGCCAGCCATATCCGAATCATGACGGCTTCAACCTCTACTACCCCAAGTTCAAC 1190
Db 1083 GTTCAGCTGGGGCCAGTCCCCGACGATGGGAACACCTTCTACTACAAACACACGAGCTT 1142
Qy 1191 AGCCACCTCTCGCCAGCGAGGCGCTCGCAATCAGTAC 1229
Db 1143 CACGTGCTGTGACGGATCATGTGGCGGCATCAATCC 1181

RESULT 7

US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862.588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2977 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus

; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 677..2776
; OTHER INFORMATION:
US-07-862-588B-1

Query Match 5.0%; Score 114.8; DB 2: Length 2977;
Best Local Similarity 57.1%; Pred. No. 3.6e-16;
Matches 250; Conservative 0; Mismatches 182; Indels 6; Gaps 2;

Qy 1382 TCCAGTACAGAACAAATGATTTCGGCGCGGGTGATAACACAGATCAAAACCGGGTCTCCAGT 1441
Db 148 TGCAGTACAGAGCGCGCGATACAAATGACGCGCAACACAGATCAAGCGGCTCTTCAACA 207
Qy 1442 TGGTGAATACGGGTGCTGCTGGTGGATTTGTGACCGGTGACGGTGCCTGTTCA 1501
Db 208 TCAAAACAAACGGTACTTCGGCTGTGATTTAAGCACGCTCAAAATCCGCTACTACTTCA 267
Qy 1502 CCCGGGATGGTGGGTGCTGACACTGGTGTACAACCTGTGACTGGGCGGCGATGGGTGTG 1561
Db 268 CCAAGGATGGTTCGCGGGGTGAACGGCTGGATC---GACTGGGCGGAGCTCGGCGGA 324
Qy 1562 GGAATATCCGCGCTCTGCGCTCGGTGAACCCCGGCGGCGGACACCTATC 1621
Db 325 GCAACATTGAGATCTCGTTTGGCAACCATACTGCGCACGAATTCGGATACGTACGTGGAGC 384
Qy 1622 TGCAGTTGCTGCTCACTGGTGGAACTTTGGCGCGCTGGTGGGTGCGAGCGGTGAGATTCAA 1681
Db 385 TGAGCTTCTGCTCGGAGGCGGCTCGATTGCGCGCGGCGGCGCAATCCGCTGAATCCAGC 444
Qy 1682 ACCGGGTGAATAAGAGTGAAGTGGTGGTCAACTTTGATGAGACCAATGACTACTCTGT---ATG 1738
Db 445 TCGCATCTCCAGAGCAGCAGCTGGTCAACTTTAAGAGCGGGAACCACTACTGTTCCGATG 504
Qy 1739 GGACGAACACCGCTTCAGGATTGCGAGAGGTGACGAGGTGTATGTCAATGCCGCGCTGG 1798
Db 505 GGACGAACACCGCTTCTGCTGACTGGGATCGGGTCTGTTGTACCAAGACGCGCAATAG 564
Qy 1799 TGTGGGGGACTGAACCGT 1816
Db 565 TGTGGGGAATGCTCCAT 582

RESULT 8

US-09-136-574A-46
; Sequence 46, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; APPLICANT: Anderson, Paige
; APPLICANT: Gibbs, Moreland
; APPLICANT: Bergquist, Peter
; APPLICANT: Daniels, Roy
; APPLICANT: Morgan, Hugh W.
; APPLICANT: Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136.574A

```
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/C1P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-136-574A-46

Query Match 3.9%; Score 90.2; DB 4; Length 2029;
Best Local Similarity 52.0%; Pred. No. 7.3e-11;
Matches 306; Conservative 0; Mismatches 268; Indels 15; Gaps 4;

Qy 1370 GTGGGTGAAGGTGACGATACAGAACAAATGATTCGGCCCGGGTGTGATAACACAGATCAAAAC 1429
Db 8 CTGGTGTCAAGGTACTGTACAGAACAAATGAGACAAGTCCGAGCACAGCTTCTATAAGGC 67

Qy 1430 CGGGTCTCCAGTTGGTGAATACGGGGTCTGTCGGTGGATTTGTTCGACGGTGAAGCGTGC 1489
Db 68 CTTGGTTTAAAGATAGTAATGAGGAGCAGCAGCAGTGTGTGATCTTAGCAGGGTTAAGATAA 127

Qy 1490 GGTACTGTTTCCACCGGAGTGTGGTCTGTCGACACTGCTGTACAACTGTCACTGGCGGG 1549
Db 128 GATAGTTGTACACAGTGGATGCTGACAAAGC---CACAGAGTGGGTATGTGACTGGGCAC 184

Qy 1550 CGATGGGTGTGGAAATATCCGCCCTGCTTCGGCTCCGTTGAACCCGGCGACGGCGAGCG 1609
Db 185 AGATAGGGGCAAGCAATGTGACATTTTGTGAAGCTTTAGCAGCGAGTGAAGTGAG 244

Qy 1610 CGGACACCTACCTCAGTTGCTTCTACTGTGGNAC-----GTTGCCCGCTGGTGGGT 1663
Db 245 CGGATTATTACTCTGAGGTAGGATTTAGCAGTGGAGCTTGGCAGCTTGCAGGCTGGTAAG 304

Qy 1664 CGAGGGTGTGAGATTCAAACCCGGGTGAATAAGAGTGAAGTGTGCAACTTTGATGAGACCA 1723
Db 305 ACACAGGGATATACAGTAAGTTTACACAGATGACTGGAGCAATTACAACTACGGCAG 364

Qy 1724 ATGACTACTCTATGGGACGACACCGGCTTCCAGGATTGGA--CGAAGGTGACGGTGT 1780
Db 365 ACGACTGGTCATGGTTGACAGAGCATGACGAATTATGGAGAGAAATCGGAAGGTGACGGTGT 424

Qy 1781 TTGTCAATGGCGGCTGGTGTGGGAGTGAACCGTCCGGCACC--AGLCCCGACCCCA 1837
Db 425 TTGTAGATGTGTTCTGTATGGGGGAGGAGCGCGGAGCGGTGACCCCAACTTCTA 484

Qy 1837 ACCCPCCCGACGCCCAACCTTCCCGGAGCGGAGCCCGACCCCAAGCGCCAGCTCCT 1897
Db 485 TCTGTATTCGTTTTCATCATTTACTCTACACCAACAGCAACGCCACACCTACACCTT 544

Qy 1895 CCGCATCCCGTCCCGGAGCGTACGCCCGACGCCCTAGCGGCTCCCGGTC 1946
Db 545 CTATCAGGATATACACGCGCCCACTCGCAACACCCACTCCGACTCCCTTC 593

RESULT 9
US-09-197-649-7/C
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Pribnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/CI-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; EARLIER FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-197-649-7

Query Match 3.4%; Score 78.4; DB 4; Length 390;
Best Local Similarity 52.8%; Pred. No. 1.8e-08;
Matches 169; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 1212 GGGCTCGCAATCAGTACGGGCACATCGCCTCCGGCGGTCCGGAGTCGAGTCCATCCCTC 1271
Db 386 GACCATCGCAACCATGGCGTCTGCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 327

Qy 1272 GTCGGCGTCTCGGTCCGCGTCTCCGTCGGCGTCTCCGTCGGCGTCTCCGTCGGCGTCTTC 1331
Db 326 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 267

Qy 1332 GTCGGCGAGCGCGTCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1391
Db 266 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 207

Qy 1392 GAACAATGATTGGCGCGCGGGGTGATTAACAGATCAACACGGGTCTCCAGTTGCTGAATAC 1451
Db 206 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 147

Qy 1452 GGGTCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1511
Db 146 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 87

Qy 1512 TGGGTCTGTCGACACTGGTGT 1531
Db 86 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 67

RESULT 10
US-09-604-913B-12
; Sequence 12, Application US/08604913B
; Patent No. 5712142
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.
; APPLICANT: Himmel, Michael E.
; APPLICANT: Baker, John O.
; APPLICANT: Chou, Yat-Chen
; TITLE OF INVENTION: METHOD FOR INCREASING
; TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: U.S.A.
```


TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2712 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-025-691-4

Query Match 2.9%; Score 65.4; DB 3; Length 2712;
Best Local Similarity 46.9%; Pred. No. 2e-05;
Matches 204; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

Qy 1845 CCCAGCCCAACCCCGTCCCGAGCCGAGCCGACCCAGCCAGCCCAAGCCCGAGCTCTCCCCATTC 1904
Db 793 CCCCAGACCCAGCCGCGCCGAGCCGCTCCAGAGCCCTTCCCGCCGCGCCG 852
Qy 1905 CCGGTCCCGAGCCGACCCGAGCCCTACGCGCTCCCGCTCCCGGGATGCTCGCGCCGC 1964
Db 853 GCTGCGCGGCTCGTCGCGCTCGCGCTCCCGCTCCCGGGATGCTCGCGCCGC 912
Qy 1965 GCGGAGTGTGCTGCTCGGTGCTGGGTGCTGGCGCGGCGGACGCTATGTGTTCAATGATTTG 2024
Db 913 CACCAGGCGCGGAGCGCGCGGCGGACCCAGCGCTCTTACACCCAGGACGCGAGCTG 972
Qy 2025 GGTCTCTGGTTTACCGCGAGCGGTGACCGTGACGATACCCGGGAGCCGCGGACGAGCGG 2084
Db 973 GGGCAGCGGCTTCGAGGCGAAGTGACCGGTGAAGAACCCGACCGCCCGCCCTCAGCGG 1032
Qy 2085 GTGACCGGTGGCGTGTCTTTGTTGGGAATCAGACGCTCAGCAACTACTTGAACACTGC 2144
Db 1033 CTGACCCCTGGAGTGGGACTTCCCGCGCGGACCAAGGTGACCTCGGCTGGAGCCCGA 1092
Qy 2145 GTTGACCAATCAGTGTGCTGCGTGACGCGGACGCAACCTGAGCTACAAACAGTGTATCA 2204
Db 1093 CGTACCAACAAACGCGACCACTGGACCGCCCAAGAACAGAGCTGGCGGGGAGCCTCGC 1152
Qy 2205 ACCGGTCACTCGACCACTTCGGATTCAACGGAAGTTACTCAGGAACAAACACCGCAC 2264
Db 1153 CCGCGGCGCTCGGTCACTTCGGCTTCAACGGCACCGGCGCCCGCCCGCTCGGGTG 1212
Qy 2265 TACACTCACTCCAC 2279
Db : 213 CAAGCTCAACGCGC 1227

RESULT 13
US-09-329-234A-6
Sequence 6, Application US/09329234A
Patent No. 6331416
GENERAL INFORMATION:
APPLICANT: Shani, Ziv
APPLICANT: Shoseyov, Oded
TITLE OF INVENTION: PROCESS OF EXPRESSING AND ISOLATING RECOMBINANT PROTEINS AND RECO
FILE OF INVENTION: PROTEIN PRODUCTS FROM PLANTS, PLANT DERIVED TISSUES OR CULTURED
FILE REFERENCE: 00/20274
CURRENT APPLICATION NUMBER: US/09/329.234A
CURRENT FILING DATE: 2001-09-04
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 1305
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Construct containing Protein L, fused to CBDcex sequence
US-09-329-234A-6

Query Match 2.8%; Score 63; DB 4; Length 1305;
Best Local Similarity 53.4%; Pred. No. 5.4e-05;

Matches 132; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
Qy 2021 ATTGGGGTCTGGGTTTACGGCGACGGTGACGCTGACGATACCGGAGCCCGGCGACGA 2080
Db 991 AGTGAACACACCGGCTTACCGCGAAGCTCACCGTGAAGAACAACGCTCTCCGCTCCGGTAG 1050
Qy 2081 GCGGGTGGACGCTGGCGTGTCTTTGGTGGGAATCAGACGCTCAGCAACTACTTGGAAACA 2140
Db 1051 ACGGCTGGACGCTCAGGTTCCCGTCCCGCCAGGAGTCAACCGGCTGGAGCT 1110
Qy 2141 CTGCGTTGACCAATCAGTGTGTCATCGGTCAGCGGACGACCTGAGCTACAAACAGTCA 2200
Db 1111 CGACGCTCAGCAGTCCGGCTCGCGCTCAGCGTCCGCAACGCCCTGCAACGCGTCA 1170
Qy 2201 TCCACCGGCTCAGTCGACCACTCTCGGATTCACCGGAGTACTCAGGAGAAACACCG 2260
Db 1171 TCCCGCGGCGGCGACCGCGGAGTTTCGGCTTCAACGGCTCGCACAGGCGACCAAGCGCG 1230
Qy 2261 CACCTAC 2267
Db 1231 CGCCGAC 1237

RESULT 14
US-08-118-200-1/c
Sequence 1, Application US/08118200
Patent No. 6197500
GENERAL INFORMATION:
APPLICANT: SUTHERLAND, Grant R
APPLICANT: RICHARDS, Robert I
APPLICANT: SCHLESSINGER, David
APPLICANT: NAGARAJA, Ramalah
APPLICANT: KREMER, Eric J
APPLICANT: YU, Sui
APPLICANT: BAKER, Elizabeth
APPLICANT: MULLEY, John C
APPLICANT: MANDEL, Jean-Louis
APPLICANT: PRITCHARD, Melanie April
APPLICANT: LYNCH, Michael
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED
TITLE OF INVENTION: FRAGILE X SYNDROME
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SNECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118.200
FILING DATE: 09-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/802,650
FILING DATE: 05-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,232
FILING DATE: 20-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/638,518
FILING DATE: 04-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/966,517
FILING DATE: 23-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113

```

: REFERENCE/DOCKET NUMBER: 020160-164
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1028 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-118-200-1

Query Match          2.7%  Score 62.4;  DB 4;  Length 1028;
Best Local Similarity 57.0%  Pred. No. 6.8e-05;
Matches 114;  Conservative 0;  Mismatches 86;  Indels 0;  Gaps 0;

Oy 1161 CGCGTTCAACGTCTACTACCAAGTTCAACAGCGACCGTCTCGCACGAGCGCGCTCGC 1220
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 553  CAGTTCACACACCGCTCTCTTTCAGCCCTGCTAGCGCGGAGCGCGCC 494
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 1221 AATCAGTACCGGCACATCGCTTCGCCGCTCGCGAGCTCGAGTCCATCTCTCGCCGTC 1280
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 493  CCGAAGAGTGGCTCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 1281 TCCGTCCGCGTCTCGGTCCGGGTCTCCGTCGCGGTCTCCGTCGCGGTCTTGTCTCGCCGAG 1340
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 433  GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 374
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 1341 CCCGTCTCCGTCTCGTCTCGC 1360
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 373  GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 15
US-08-458-745-1/c
: Sequence 1, Application US/08458745
: Patent No. 6242576
: GENERAL INFORMATION:
: APPLICANT: SUTHERLAND, Grant R
: APPLICANT: RICHARDS, Robert I
: APPLICANT: SCHLESSINGER, David
: APPLICANT: NAGARAJA, Ramaiah
: APPLICANT: KREMER, Eric J
: APPLICANT: YU, Sui
: APPLICANT: BAKER, Elizabeth
: APPLICANT: MULLEY, John C
: APPLICANT: MANDEL, Jean-Louis
: APPLICANT: PRITCHARD, Melanie April
: APPLICANT: LYNCH, Michael
: TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED
: TITLE OF INVENTION: FRAGILE X SYNDROME
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,745
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/118,200
: FILING DATE: 09-SEP-1993
: APPLICATION NUMBER: US 07/802,650
```

```

: FILING DATE: 05-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/672,232
: FILING DATE: 20-MAR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/638,518
: FILING DATE: 04-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/966,517
: FILING DATE: 23-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Crane-Feury, Sharon E
: REGISTRATION NUMBER: 36,113
: REFERENCE/DOCKET NUMBER: 020160-164
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1028 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-458-745-1

Query Match          2.7%  Score 62.4;  DB 4;  Length 1028;
Best Local Similarity 57.0%  Pred. No. 6.8e-05;
Matches 114;  Conservative 0;  Mismatches 86;  Indels 0;  Gaps 0;

Oy 1161 CGCGTTCAACGTCTACTACCAAGTTCAACAGCGACCGTCTCGCACGAGCGCGCTCGC 1220
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 553  CAGTTCACACACCGCTCTCTTTCAGCCCTGCTAGCGCGGAGCGCGCC 494
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 1221 AATCAGTACCGGCACATCGCTTCGCCGCTCGCGAGCTCGAGTCCATCTCTCGCCGTC 1280
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 493  CCGAAGAGTGGCTCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 1281 TCCGTCCGCGTCTCGGTCCGGGTCTCCGTCGCGGTCTCCGTCGCGGTCTTGTCTCGCCGAG 1340
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 433  GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 374
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 1341 CCCGTCTCCGTCTCGTCTCGC 1360
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 373  GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Search completed: November 13, 2002, 10:50:38
Job time : 118 secs
```


; PRIOR FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: 07/829,461
; PRIOR FILING DATE: 1992-01-31
; PRIOR APPLICATION NUMBER: 07/739,055
; PRIOR FILING DATE: 1991-08-01
; PRIOR APPLICATION NUMBER: 07/561,968
; PRIOR FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-790-399-7

Query Match 3.4%; Score 78.4; DB 10; Length 390;
Best Local Similarity 52.8%; Pred. No. 1.3e-09;
Matches 169; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 1212 GCCGCTCCCAATCAGTACCGGCACATCCCTCCGCGTGGCGGAGCTCGAGTCCATCCTC 1271
Db 386 GACCATCCCAACCATGGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 327
Qy 1272 GTCGCCGCTCCGTCGGCGCTCTCGTGGCGCTCTCCGTCGGCGCTCTCGTCGGCGCTTC 1331
Db 326 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 267
Qy 1332 GTCGCCGAGCCGCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1391
Db 266 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 207
Qy 1392 CAACAATGATTCGGCGCGGGGTATACCAACACATCAACCGGCTCCAGATTGGTGAATAC 1451
Db 206 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 147
Qy 1452 GGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1511
Db 146 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 87
Qy 1512 TGGTCGTCGACATCGTGT 1531
Db 86 GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 67

RESULT 5
US-09-974-300-684
; Sequence 684, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 684
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-684

Query Match 3.0%; Score 68.8; DB 10; Length 1314;

Best Local Similarity 52.9%; Pred. No. 3.4e-07;
Matches 148; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 1535 ACTGTGACTGGCGCGGATGGGTGTGGGAATATCCGGCGCTCTGTTCCGGCTCGGTGAACC 1594
Db 1032 ACTGGACTATGCCCAATCGGCTGCAGCAAAATCAGGCACAAATTCGTTCAATTAATAA 1091
Qy 1595 CGGCGACCGCGGCGGACACCTACCTGCAGTTCGTTCTGCTTCACTGCTGCAAGCTTGGCGG 1654
Db 1092 AAGCGTTAAACGAGCAGACAGCTATCTTGAAGTAGGATTTAAAAATGGTACATTGGCGC 1151
Qy 1655 CTGGTGGGTGCGACGGTTCAGATTCAAAAACCGGTCGAATAAGAGTGAAGTTCGTCGAACCTTTG 1714
Db 1152 CGGGGCTCATACTGCGCAAAATCCAGATCCGCTTCACAATCAGCGCTGGAGCAATTATG 1211
Qy 1715 ATGAGACCAATGACTACTGCTATGGGAGCAACACCGCTTCCAGGATTGGAGCAAGTGA 1774
Db 1212 CCCAAGCGGCGACTATTTCAATTTTAAATTCAAACACGCTTTAAAAATACGAAAAAATCA 1271
Qy 1775 CGGTGTATGCAATGCGCGCTGTTGTCGGGAGCTGAACC 1814
Db 1272 GTTGTATGAGAACGGAAGCTGATTTGGGCACTGAACC 1311

RESULT 6
US-09-748-033-4
; Sequence 4, Application US/09748033
; Patent No. US2002069431A1
; GENERAL INFORMATION:
; APPLICANT: Broadway, Roxanne M.
; APPLICANT: Gongora, Carmenza E.
; TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOIDASE AND THEIR
; TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
; FILE REFERENCE: 19603/3091
; CURRENT APPLICATION NUMBER: US/09/748,033
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/172,003
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2712
; TYPE: DNA
; ORGANISM: Streptomyces albidoflavus
US-09-748-033-4

Query Match 2.9%; Score 65.4; DB 10; Length 2712;
Best Local Similarity 46.9%; Pred. No. 2.6e-06;
Matches 204; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

Qy 1845 CCCAGGCCCAACCCCGTCCCGAGCGCCGAGCCGAGCCCAAGCCCAAGCCCAAGCTCTCCCCATC 1904
Db 793 CCCAGCACCCGAGCGCGCCGAGCGCTCCAGACCCCTCAGCGCTTCCGCGCGCGCGC 852
Qy 1905 CCCGTCCCGAGCCCGAGCCCGCCAGCCCTTACCGCGCTCCCGCTCGCGAGCCCGCTCCCGCTC 1964
Db 853 GCTGCGCGGCTCGTCGCGCTCGCGGTCCTTCCGCGGATGTCGCGCTCGCGCGCCC 912
Qy 1965 GCCAGTGTGTCGTCGCGGTCGCGGTCGCGGCGACGATATGTGTGAATAGTATTG 2024
Db 913 CACCCAGCGCGCGAGCGCGCGCGCCAGCCCGCCAGCGCTTCTACACAGCAGCAGGACTG 972
Qy 2025 GGGTTCGGGTTTACGGCGAGGTGACGGTACCAATACCGGAGCGCGGAGCGAGCGG 2084
Db 973 GGGCAGCGGCTTCAGGGCAAGTGGAGCGGTGAAGAACAACCGCGCCCGCCCTCAGCGG 1032
Qy 2085 GTGACGGTGGCGTGGTCTGTTGTTGGGAATCAGACGGTACACGAATCTACTGAACACTGC 2144
Db 1033 CTGGACCTGGAGTGGGACTTCCCGCGCGGNAACCAAGGTGACCTCGGCTGGGAGCGGA 1092
Qy 2145 GTTGACCAATCAGGTGCATCGGTGACGGGAGCAACCTGAGCTTACAAACAGTGTATCCA 2204
Db 1093 CGTACCACCAACCGCGGACCACTGGACCGCCGCAAGCAAGAGCTGGGCGGAGCCCTCGC 1152


```
RESULT 10
US-09-974-300-3603
; Sequence 3603, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3603
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-3603

Query Match      2.5%; Score 58.2; DB 10; Length 300;
Best Local Similarity 55.1%; Pred. No. 6.6e-05;
Matches 158; Conservative 0; Mismatches 123; Indels 6; Gaps 2;

Qy 935 ACTGGGCCAGACCGCGACTGGGGAACGCAATGGATCAAGCATATATTCGGATGCGG 994
Db 3 ATTGGGGTTATACGAGGCGAGTGGGGGATTCAGTGATCACAGACGAGATTTGCGATGGGA 62

Qy 995 CAGCGATCGGCAAGCCGACCAATTCGGAAGAAATTCGGCTGGCAGACACCGGACCGC---G 1051
Db 63 AAGNAATCGGCAAGCCGTCGTTTATAGAGAGATCCGGCTATCAGTATAGTTCGAGAGGG 122

Qy 1052 ATTCGCTCTATCAGACGTGGACCGACTGTGCGGTACGAAAGCGGTGAAGAGCGTGGAACT 1111
Db 123 ACTAGCTCTACAGAACCTGCTGTTCAACTCATAGAAAGCAGAGCGGTGCGGGCAGCCAAAT 182

Qy 1112 TCTGATGCTCGCTGGGA---ATCTCAAGCCGCGCCATATCCGAACTATACGCGCTTCA 1168
Db 183 TTTGGATTTTGACCGGATTCAGATGACGCGGACCCCTTTATCCGGACTATACCGGTTTC 242

Qy 1169 ACGTCTACTACCCAAAGTTCAACAGCAGCGCTCCCTCGCGCAGGAGCGG 1215
Db 243 GGATCGTTTATCCGAGCTCTCCCGCTTCGTCATTTTCAGACGACGCG 289

RESULT 11
US-09-860-846-34/c
; Sequence 34, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-34

Query Match      2.5%; Score 57.4; DB 10; Length 4689;
Best Local Similarity 46.4%; Pred. No. 0.00023;
Matches 187; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 1209 CGAGGCGCTCGCAATCAGTACCGGCGACATCGCCCTCGCGCGTCCGAGTCCATC 1268
Db 3912 CGCGTCGTGGCGCGCGCCGATCGCGTCCGCGCCAGAGGCCGCCAGCCGAGGT 3853

Qy 1269 CTCGTCGCGCGTCTCGGTGCGCGTCTCCGTCCGCGGTCTCCGTCCGCGGTTC 1328
Db 3852 CGCGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3793

Qy 1329 TTCGTGCGCGGAGCGCGTCTCGGTCTCGGTCTCGGTCTCGGTCTCGGTCTCGGT 1388
Db 3792 CGCGTAGAGCGCGTTCGGCTGCGCGTCCCGCAGACCGCGGTTCGAGGAGTAGAGACGAA 3733

Qy 1389 CAAGAACAATGATTTCGGCGCGCGGTGATAACAGATCAACACGGGTCTCCAGTTGGTCAA 1448
Db 3732 GCGCTCCAGCGGAGTGGCGCGCGAGCAGGTCTGTCGAGGACCTCGCGCGCGCGTCTGTCG 3673

Qy 1449 TACGGGTCGTCTCGGTGGATTTCCTCCACGCTGACGCTGCGGTACTGTTCCACCGGGA 1508
Db 3672 GCCAGGATCGCGCGATGTCCTCGCGCGCGGTGACGTCCAGCGGATCGCGCGCGGTGCG 3613

Qy 1509 TCGTGGTCTCGACACTGGTGTACAACGTGCTACTGGCGCGCGATGGGTGTGGCAATAT 1568
Db 3612 CGCGCGGTGTGGAGCAGCGCGGTGACGGCGTCTCGCGCGGATGCGGTTCGAGGAGGT 3553

Qy 1569 CGGCGCTCTGTTGCGGTCTGATGACCGCGCGGCGCGCGCGCGCGCGCGCGCG 1611
Db 3552 CGCATGCGGTGGGGTTCGCGCGACGTCCGACGCGCGCGATGGTG 3510
```

```
RESULT 12
US-09-861-289-34/c
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-34

Query Match      2.5%; Score 57.4; DB 10; Length 4689;
Best Local Similarity 46.4%; Pred. No. 0.00023;
Matches 187; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 1209 CGAGGCGCTCGCAATCAGTACCGGCGACATCGCCCTCGCGCGTCCGAGTCCATC 1268
Db 3912 CGCGTCGTGGCGCGCGCCGATCGCGTCCGCGCCAGAGGCCGCCAGCCGAGGT 3853

Qy 1269 CTCGTCGCGCGTCTCGGTGCGCGTCTCCGTCCGCGGTCTCCGTCCGCGGTTC 1328
Db 3852 CGCGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3793

Qy 1329 TTCGTGCGCGGAGCGCGTCTCGGTCTCGGTCTCGGTCTCGGTCTCGGTCTCGGT 1388
Db 3792 CGCGTAGAGCGCGTTCGGCTGCGCGTCCCGCAGACCGCGGTTCGAGGAGTAGAGACGAA 3733

Qy 1389 CAAGAACAATGATTTCGGCGCGCGGTGATAACAGATCAACACGGGTCTCCAGTTGGTCAA 1448
Db 3732 GCGCTCCAGCGGAGTGGCGCGCGAGCAGGTCTGTCGAGGACCTCGCGCGCGCGTCTGTCG 3673

Qy 1449 TACGGGTCGTCTCGGTGGATTTCCTCCACGCTGACGCTGCGGTACTGTTCCACCGGGA 1508
Db 3672 GCCAGGATCGCGCGATGTCCTCGCGCGCGGTGACGTCCAGCGGATCGCGCGCGGTGCG 3613

Qy 1509 TCGTGGTCTCGACACTGGTGTACAACGTGCTACTGGCGCGCGATGGGTGTGGCAATAT 1568
Db 3612 CGCGCGGTGTGGAGCAGCGCGGTGACGGCGTCTCGCGCGGATGCGGTTCGAGGAGGT 3553

Qy 1569 CGGCGCTCTGTTGCGGTCTGATGACCGCGCGGCGCGCGCGCGCGCGCGCGCGCG 1611
Db 3552 CGCATGCGGTGGGGTTCGCGCGACGTCCGACGCGCGCGATGGTG 3510
```



```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: cDNA clone of DYT1 Intron 1
; NAME/KEY: misc_feature
; LOCATION: (1)..(283)
; OTHER INFORMATION: n = A,T,C or G
US-09-772-105-48

Query Match      2.4%; Score 55; DB 10; Length 283;
Best Local Similarity 63.0%; Pred. No. 0.00036;
Matches 85; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 1827 CCCAGCCCCACACCCAGCCCGCCGAGCCCAACCCCGTCCCGGAGCCCGGACCCCAAG 1886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 CCCACCCCGCAGCCCGCCGCGCGTCCATCCCGCAGCCCGCCAGCCCAACCCGCC 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1887 CCCAGGTCTCTCCCGATCCCGGTCGCCGAGCCCGCCAGCCCGCCAGCCCTACGGCGTCCCGGTC 1946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 CTAGTGCCATCGCCGAGCCCGCCAGCCCGCCAGCCCGCCGCTCCAGCCCTAGTCTACGCCCGGC 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1947 GCCGAGCCCGTCGCC 1961
    ||| ||| |||
Db 118 CCTGCTGCCATCCCC 104
    ||| ||| |||

Search completed: November 13, 2002, 11:49:40
Job time : 179 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:49:26 ; Search time 33.6746 Seconds
(without alignments)
1483.879 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTASGQFVNLGLPY.....YPSSTATVLASEALAISTG 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829.5	39.9	666	19 AAW34992	Thermotoga neapoli
2	812	39.1	663	19 AAW49872	Thermotoga maritim
3	812	39.1	680	19 AAW35005	Thermotoga maritim
4	806	38.8	680	18 AAW34564	Thermotoga maritim
5	805	38.7	680	19 AAW49868	Thermotoga maritim
6	543.5	26.1	431	21 AAG27501	Arabidopsis thalia
7	543.5	26.1	431	21 AAG42172	Arabidopsis thalia
8	543.5	26.1	431	23 ABB93972	Herbicidally activ
9	543.5	26.1	442	21 AAG42171	Arabidopsis thalia
10	543.5	26.1	443	21 AAG27500	Arabidopsis thalia

11	537	25.8	408	23 ABB92275	Herbicidally activ
12	532.5	25.6	416	22 AAG79297	Amino acid sequenc
13	517.5	24.9	431	23 ABB90809	Herbicidally activ
14	516	24.8	414	23 ABB92274	Herbicidally activ
15	498	24.0	448	23 ABB93325	Herbicidally activ
16	495	23.8	431	23 ABB93133	Herbicidally activ
17	493	23.7	427	21 AAY93441	Amino acid sequenc
18	477	22.9	375	21 AAG27502	Arabidopsis thalia
19	477	22.9	375	21 AAG42173	Arabidopsis thalia
20	468	22.5	351	23 ABB92490	Herbicidally activ
21	459	22.1	415	22 AAE00417	Lycoposicon escul
22	426.5	20.5	369	20 AAY14150	Protein encoded by
23	411.5	19.8	403	23 ABB91820	Herbicidally activ
24	386	18.6	312	21 AAG07279	Arabidopsis thalia
25	347.5	16.7	245	21 AAG07280	Arabidopsis thalia
26	346.5	16.7	230	21 AAG07281	Arabidopsis thalia
27	323	15.5	264	21 AAG47479	Arabidopsis thalia
28	323	15.5	318	21 AAG47478	Arabidopsis thalia
29	315	15.2	264	21 AAG22457	Arabidopsis thalia
30	315	15.2	318	21 AAG22456	Arabidopsis thalia
31	261	12.6	236	21 AAG30018	Arabidopsis thalia
32	261	12.6	236	21 AAG22458	Arabidopsis thalia
33	261	12.6	236	21 AAG47480	Arabidopsis thalia
34	238	11.4	231	21 AAG20019	Arabidopsis thalia
35	201	9.7	210	21 AAG20020	Arabidopsis thalia
36	181.5	8.7	521	22 AAB48786	Acidothermus cellu
37	181.5	8.7	562	21 AAY69508	Acidothermus cellu
38	179.5	8.6	358	19 AAW39262	A. cellulolyticus
39	179.5	8.6	521	17 AAR89927	A. cellulolyticus
40	175.5	8.6	562	23 AAW79549	A. cellulolyticus
41	175.5	8.4	521	22 AAB48787	Acidothermus cellu
42	173.5	8.3	521	22 AAB48788	Acidothermus cellu
43	134	6.4	1684	12 AAR14948	Bacterial amylase
44	127	6.1	425	22 ABB09060	Thermus caldophilu
45	124.5	6.0	711	19 AAW55103	Streptococcus pneu

ALIGNMENTS

RESULT 1

AAW34992

ID AAW34992 standard; Protein: 666 AA.

AC AAW34992;

XX 21-MAY-1998 (first entry)

DT Thermotoga neapolitana endoglucanase.

DE Endoglucanase; cellulase; carboxymethylcellulose; cellulose;

KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;

KW thermostable enzyme; thermophilic; glycosidase.

XX Thermotoga neapolitana (Clone 56GP1).

XX WO9744361-A1.

PD 27-NOV-1997.

XX 22-MAY-1997; 97WO-US08793.

XX 22-MAY-1996; 96US-0651572.

XX (RECO-) RECOMBINANT BIOCATALYSIS INC.

PI Lam DE, Mathur EJ;

XX WPI: 1998-018435/02.

DR N-PSDB; AAT94200.

PT Endoglucanase(s), preferably form archaeal bacterium, AEP11 la -
useful to degrade carboxymethylcellulose and hydrolyse of

```
PT  beta-1,4-glycosidic bonds in cellulose
XX
PS  Claim 1: Fig 1H: 164pp; English.
XX
CC  This protein comprises an endoglucanase of Thermotoga neapolitana
CC  (Clone 56GP1) that is capable of degrading carboxymethylcellulose
CC  and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It
CC  has homology to an endoglucanase of archaeobacterium AEP11a (see
CC  AAW34985). It can be produced from native cells or from recombinant
CC  host cells, especially prokaryotic host cells transformed with a
CC  plasmid or virus-derived vector including the endoglucanase DNA
CC  (see AAT94200). 24 Endoglucanases (see AAW34986-W35008) are claimed.
CC  They can be used to degrade cellulose for the conversion of plant
CC  biomass into fuels and chemicals, for use in detergents, textiles,
CC  animal feed, waste treatment, and in the fruit juice and brewing
CC  industries for the clarification and extraction of juices.
XX
SQ  Sequence 666 AA:
      Query Match      39.9%; Score 829.5; DB 19; Length 666;
      Best Local Similarity 43.5%; Pred. No. 4e-60;
      Matches 173; Conservative 63; Mismatches 119; Indels 43; Gaps 13;
QY  1  APAGFVTASGGQFVLNGLPYRYGGTNNYYLSYQSHADVDVLAQAQAMNLSVIRTWGFI 60
DB  1  ANSDFVKVNGRFTLNGEERFVGSNNYYMHYKSNRMIDSVLESAMKGVKVLRIWGF-- 76
QY  61  IGSIDGVSPTIDGNKNGFYQYWDPTGA---PAYNDGPTGLOGLDYAIAASAAHGLRVI 117
DB  77  ---LDGESYCRDKN-----TYMHPAPGVFCLPGCTNAQDGERLDYTVAKAKELGKLI 127
QY  118  VVLTDNWKKEFGMDQYDKWYGLPYHDNFYTDPRTOQAYKNVNNHLLNRVNSITGVYKND 177
DB  128  IVLVNNMDDFGGMQYVRFWFGIHHDFYRNEKIKEEYKVVSLINRVNTYTCVPYREE 187
QY  178  PTIFAWELANEPRCVSGSGLTPTSGCTQATIVNNVDQMSAYKSIDPNHMYSVGDGEF-- 235
DB  188  PTIMAWELANEPRC-----ETDKSGN-----TLVVEWVEMSAYKISLDPNHLVAVGDEGFN 239
QY  236  -YIGSTGSG---WPYNDPDCGVDDNALLRVKNIIDFGTYHLYPNYWGON----ADWGTOW 287
DB  240  NYEGFRPYGGEAWAYNGWS-GVDWKRLLLEIETVDFGTFHLYPSHWGVSPEYNAQWGA 298
QY  288  IKDHIAAAIGKPTILEEFG--WQTP-DRDSVYOTVTTQVTRTNGEAGWNFMLAGNVNG 344
DB  299  IEDHIKIAKEVGKPVLEEVGIPKSAVNRYALVKNLDVYLNLCGNGAMFMMLAGIGEG 358
QY  345  QP-----YPNYDGFNVYPSSTATVLAASE-ALAI 375
DB  359  WDRDEKGYPDYDGFRIVNDSEAKLIREYAKLFSTG 396
RESULT 2
AAW49872
ID  AAW49872 standard; Protein: 563 AA.
XX
AC  AAW49872:
XX
DT  21-DEC-1998 (first entry)
XX
DE  Thermotoga maritima MSB8-6GP2 glycosidase.
XX
KW  Glycosidase; MSB8-6GP2; thermostable enzyme; oligosaccharide;
XX  glucose; sugar; baking; textile; detergent; beta-galactosidase.
XX
OS  Thermotoga maritima strain MSB8-6GP2.
XX
PN  WO9824799-A1.
XX
PD  11-JUN-1998.
XX
PF  08-DEC-1997; 97WO-US22623.
XX
```

```
PR  10-OCT-1997; 97US-0949026.
XX  06-DEC-1996; 96US-0056916.
XX  (DIVE-) DIVERSA CORP.
XX
PI  Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
XX
DR  WPI: 1998-362407/31.
DR  N-PSDB: AAW36921.
XX
XX  Glycosidase enzymes from organisms of the genera Staphylothermus,
PT  Pyrococcus and Thermococcus - for deriving sugar from
PT  oligosaccharides, useful in the e.g. food processing, textile or
XX  baking industries
XX
XX  Claim 1: Fig 15a-d; 92pp; English.
XX
CC  This is the amino acid sequence of glycosidase MSB8-6GP2, deduced
CC  from a polynucleotide (see AAW36921) of clone 6GP2 of Thermotoga
CC  maritima MSB8 that grows optimally at 85 degC in high salt medium.
CC  The invention provides 18 polynucleotides (see AAW36907-24) coding
CC  for thermostable glycosidases (see AAW49858-75) having glucosidase,
CC  alpha-galactosidase, beta-galactosidase, beta-mannosidase,
CC  beta-mannanase, endoglucanase or pullulanase activity. Vectors and
CC  host cells are also claimed. A method is provided for producing
CC  the enzymes by recombinant techniques. A claimed method for
CC  generating glucose from soluble cell oligosaccharides comprises
CC  contacting a sample (selected from dairy products, fruit juice,
CC  detergent, textile, guar gum, animal feed, plant biomass or waste
CC  product) containing oligosaccharides (selected from maltose,
CC  cellobiose, lactose, sucrose, raffinose, stachyose, verbascose,
CC  cellulose, starch, amylose, glycogen, disaccharides, polysaccharides
CC  and pullulan) with one of the claimed glycosidases such that glucose
CC  is produced.
XX
SQ  Sequence 663 AA:
      Query Match      39.1%; Score 812; DB 19; Length 663;
      Best Local Similarity 42.8%; Pred. No. 1.1e-58;
      Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;
QY  5  FVTASGGQFVLNGLPYRYGGTNNYYLSYQSHADVDVLAQAQAMNLSVIRTWGFI 64
DB  17  FVKVENGKFKALNGKEFRFIGSNYYMHYKSNMIDSVLESARDMGIKVLRIWGF----L 71
QY  65  DGSVPTIDGNKNGFYQYWDPTGAPYNDG----PTGLOGLDYAIAASAAHGLRVI 120
DB  72  DGSYCRDKN-----TYMHPAPGVFVPEGISNAQSGFERLDYTVAKAKELGKLI 125
QY  121  TNDWKEFGMDQYDKWYGLPYHDNFYTDPRTOQAYKNVNNHLLNRVNSITGVYKNDPT 180
DB  126  VNNMDDFGGMQYVRFWFGTTHDDFYRDEKIKEEYKVVSLVNLHVNTYTCVPYREEPT 185
QY  181  FAWELANEPRCVSGSGLTPTSGCTQATIVNNVDQMSAYKSIDPNHMYSVGDGEF---Y 237
DB  186  MAWELANEPRC-----ETDKSGN-----TLVVEWVEKSSYKISLDPNHLVAVGDEGFSNYE 237
QY  238  GSTGSG---WPYNDPDCGVDDNALLRVKNIIDFGTYHLYPNYWGON----ADWGTOW 290
DB  238  GFKPYGGEAWAYNGWS-GVDWKRLLLEIETVDFGTFHLYPSHWGVSPEYNAQWKA 296
QY  291  HIAAAIGKPTILEEFG--WQTP-DRDSVYOTVTTQVTRTNGEAGWNFMLAGNVNGOP- 346
DB  297  HIKAKEIGKPVLEEVGIPKSAVNRYALVKNLDVYLNLCGNGAMFMMLAGIGEGSDR 356
QY  347  -----YPNYDGFNVYPSSTATVLAASE-ALAI 375
DB  357  DERGGYPDYDGFRIVNDSDSEAKLIREYAKLFNTG 391
RESULT 3
AAW35005
ID  AAW35005 standard; Protein: 680 AA.
```



```

XX AC AAW35005;
XX DT 21-MAY-1998 (first entry)
XX DE Thermotoga maritima endoglucanase.
XX KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
XX KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
XX KW thermostable enzyme; thermophilic; glycosidase.
XX OS Thermotoga maritima strain MSB8 (Clone 6P2).
XX PN WO9744361-A1.
XX PD 27-NOV-1997.
XX PF 22-MAY-1997; 97WO-US08793.
XX PR 22-MAY-1996; 96US-0651572.
XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX PI Lam DE, Mathur EJ;
XX XX WPI; 1998-018435/02.
XX DR N-PSDB; AAT94213.
XX PT Endoglucanase(s), preferably form archaeal bacterium, AEP11 la -
XX PT useful to degrade carboxymethylcellulose and hydrolyse of
XX PT beta-1,4-glycosidic bonds in cellulose
XX PS Claim 1; Fig 1U; 164pp; English.
XX CC This protein comprises an endoglucanase of Thermotoga maritima MSB8
XX CC (Clone 6P2) that is capable of degrading carboxymethylcellulose
XX CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It
XX CC has homology to an endoglucanase of archaebacterium AEP11a (see
XX CC AAW34985). It can be produced from native cells or from recombinant
XX CC host cells, especially prokaryotic host cells transformed with a
XX CC plasmid or virus-derived vector including the endoglucanase DNA
XX CC (see AAT94213). 24 Endoglucanases (see AAW34986-W35008) are claimed.
XX CC They can be used to degrade cellulose for the conversion of plant
XX CC biomass into fuels and chemicals, for use in detergents, textiles,
XX CC animal feed, waste treatment, and in the fruit juice and brewing
XX CC industries for the clarification and extraction of juices.
XX SQ Sequence 680 AA;

Query Match 39.1%; Score 812; DB 19; Length 680;
Best Local Similarity 42.8%; Pred. No. 1.2e-58;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

OY 5 FVTASGGOFLVGLPYRYGCTNNYYLSYQSHADVDVLAQAAMNLSVIRTWGFDIGSL 64
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34 FVKVNGKFKALNGKEFRFGISNNYYMHYKSGMIDSVLESARDMGIKVLRIWGF-----L 88
OY 65 DGSVPTIDGNKNGFYFOYWDPTGAPAYNDG----PTGLOGLDYAIASAAHGLRVIVL 120
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
89 DGESYCRDKN-----TYMHPGCVGVPFGISNAQSGFERLDYTVAKAKELGKLVIVL 142
OY 121 TNDWKEFGGMDQYKWKYGLPYHNFYDTPRTOQAYKKNVHLLNRVNSITGVTKNDPTI 180
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
143 VNNMDDFGGNQYRVFGGTHDDFYRDEIKKEYKYVSVFLVNHVNTYTGVPYREEPTI 202
OY 181 FANELANERPCVSGTLPTSGTCTQATIVANNVDOMSAYVKSID .JVSYGDSGF---YI 237
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 MANELANERPC-----ETDKSGN-----TLVEWVKEMSSYIKSLDPNHLVAVYDGEFFSNE 254
OY 238 GSTQGSQ---WPYNDPSGDVNNALLRKVNIDFGTYHLYPNYWGON----ADWGTOWIKD 290
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255 GPKYPGGEAEWYNGWS-GVDWKKLLSIETVDFGTFLYPHSHWGVSPENYAQNGAKWIED 313

```

```

OY 291 HIANAAIGKPTILEFG--WQTP--DRDSVYOTWTOTVRTNCEAGNFMMLAGNVNGOP- 346
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
314 HIKIAKEIGKPVLEEYGIPIKSAVNTAIYRLWMDLVYDLGGDGAMFMMLAGIGGSDK 373
OY 347 -----YPNYDGFNVYSPSTATVLAASE-ALAISTG 375
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 DERGYYPDYDGFRIVNDSDSPEAELIREYAKLFNTG 408
RESULT 4
ID AAW34564 standard; Protein; 680 AA.
XX AC AAW34564;
XX DT 12-MAR-1998 (first entry)
XX DE Thermotoga maritima beta-mannanase.
XX KW Glycosidase; thermostable; textile; food processing; pharmaceutical;
XX KW detergent; baking; industry; Thermococcus; Staphylothermus;
XX KW Pyrococcus; glucose; soluble oligosaccharide; beta-mannanase.
XX OS Thermotoga maritima.
XX PN WO9725417-A1.
XX PD 17-JUL-1997.
XX PF 10-JAN-1997; 97WO-US00092.
XX PR 13-SEP-1996; 96US-0712612.
XX PR 11-JAN-1996; 96US-0583787.
XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
XX WPI; 1997-372858/34.
XX DR N-PSDB; AAT93686.
XX CC New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
XX CC PT and Pyrococcus, used in the textile, food processing,
XX CC PT pharmaceutical, detergent and baking industries
XX PS Claim 4; Fig 11; 82pp; English.
XX CC The present sequence represents beta-mannanase isolated from Thermotoga
XX CC maritima. The enzyme or its encoding nucleic acid sequence is
XX CC used for generating glucose from soluble oligosaccharides. The enzyme
XX CC can be used in the food processing, pharmaceutical, textile, detergent
XX CC and baking industries. The enzyme is also used to treat lactose
XX CC intolerance, as a diagnostic reporter molecule, in corn wet milling or
XX CC in the fruit juice industry. The enzymes can be used to hydrolyse guar
XX CC gum to remove non-reducing terminal mannose residues. The nucleic acids
XX CC encoding the enzyme may be used to generate probes to identify similar
XX CC sequences.
XX SQ Sequence 680 AA;

Query Match 38.8%; Score 806; DB 18; Length 680;
Best Local Similarity 42.3%; Pred. No. 3.6e-58;
Matches 167; Conservative 62; Mismatches 122; Indels 44; Gaps 13;

OY 5 FVTASGGOFLVGLPYRYGCTNNYYLSYQSHADVDVLAQAAMNLSVIRTWGFDIGSL 64
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34 FVKVNGKFKALNGKEFRFGISNNYYMHYKSGMIDSVLSDPRDNGIKVLRIWGF-----L 88
OY 65 DGSVPTIDGNKNGFYFOYWDPTGAPAYNDG----PTGLOGLDYAIASAAHGLRVIVL 120
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
89 EGESYCRDKN-----TYMHPGCVGVPFGISNAQSGFERLDYTVAKAKELGKLVIVL 142
OY 121 TNDWKEFGGMDQYKWKYGLPYHNFYDTPRTOQAYKKNVHLLNRVNSITGVTKNDPTI 180

```


PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142134.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 23-APR-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160988.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
<div>Query Match 26.1%; Score 543.5; DB 21; Length 431; Best Local Similarity 35.5%; Pred. No. 1.le-36; Matches 139; Conservative 49; Mismatches 139; Indels 65; Gaps 17;</div>			
QY	4	GFVTASGGQFVLNLPYRGFTNNTLSY-----QSHADVDVLAQAQAMNLSVIRTWGF	58
DB	30	GFVRTKGVOFSLNGPYVANGFNAYWLMYVASDPSORSKISTAFODASRHGLTVARTWAF	89
QY	59	IDIGSLDSVPTIDGNKNGFFQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRVIV	118
DB	90	SD-----GGYRALQYSPG-----SYNEDM--FQGLDFALAEARRHGKIL	128
QY	119	VLTDNWKFEFGMDQYDKW---YGLPY--HDNFYDPTQQAQYKKNVNLNRVNSITGYT	173
DB	129	SFANNYESFGGRQYVDNARSGRPVSSDEDFDLSLVKDFYKNIKAVLNRFNFTTKVH	188
QY	174	YKNDPTIFAWELANPRCVGSCTLPTSGTCTQATIVNVVDQMSAYVKSIDPNHMYSVGDE	233
DB	189	YKDDPTIMAWELMNEPRCPD-----PSGRAIQA-----WITEMAAHVKSILDRNHLLEAGLE	240
QY	234	GFYIGSTOGSGWPNYNDPSD-GVDNNALLRVKNIDPCTYHLVPNYWGQNA-----DWCTQ	286
DB	241	GFY-GQSSPQSKTLNPPQGFQDFIANNRIPGIDFVTVHSYPDEWFPDSSQSQMDFLNK	299
QY	287	WIKDHIAAA-AIGKPTTILEEF-----GWQTPDPRDSVYOT-----WTQTVRTNGEAGW	333
DB	300	WLDIAHQDAQNVLHKPIILAEFGKSMKKPGYTPAQRDIVFNTVYSKIYGSAKRGCAAAAGG	359
QY	334	NFWMLAGNVNGQYPNY-DGFNVVYPSSTATV	364
DB	360	LFWQLL--VNG--IDNFQDGYGIIILSQSSSTV	387
RESULT 7			
ID	AAG42172 standard; Protein; 431 AA.		
XX	AC	AAG42172;	
XX	DT	18-OCT-2000 (first entry)	
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 52560.	
KW	KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
OS	OS	termination sequence.	
XX	XX	Arabidopsis thaliana.	
PN	PN	EPI033405-A2.	

XX	06-SEP-2000.	2000EP-0301439.
PD		
XX		
PF	25-FEB-2000;	99US-0121825.
XX		99US-012180.
PR	25-FEB-1999;	99US-0123180.
PR	05-MAR-1999;	99US-0123548.
PR	09-MAR-1999;	99US-0125788.
PR	23-MAR-1999;	99US-0126264.
PR	25-MAR-1999;	99US-0126785.
PR	29-MAR-1999;	99US-0127462.
PR	01-APR-1999;	99US-0128234.
PR	06-APR-1999;	99US-0128714.
PR	08-APR-1999;	99US-0129845.
PR	16-APR-1999;	99US-0130077.
PR	19-APR-1999;	99US-0130449.
PR	21-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.

```
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 26.1%; Score 543.5; DB 21; Length 431;
Best Local Similarity 35.5%; Pred. No. 1.le-36;
Matches 139; Conservative 49; Mismatches 139; Indels 65; Gaps 17;

Qy 4 GFVTASGCOFVNLGPLYRGCTNNYLSY-----QSHADVDVLAQAAMNLSVIRTWCF 58
Db 30 GFVRTKGVOFSLNGPYFYANGFNAYWLMYVASDPSQRSKISTAFQDASRHLTVARTWAF 89
Qy 59 IDIGSLDGSVPTIDGNKNGFYQYWDPSGTGAPAYNDGPTGLQGLDYATASAAHGLRVIV 118
Db 90 SD-----CGYRALQYSPG-----SYNEDM--FQGLDFALAEARRHGIKIL 128
Qy 119 VLTNDWKEFGGMDYDKW---YGLPY--HDNFTDPRTOQAYKNVYNHLLNRVNSITGVT 173
Db 129 SFANNESFGGRKQYVDWARSGRPRVSSDEDDFTDSLKVDFYKNHIIKAVLNRFTTKVH 188
Qy 174 YKNDPTIFAWELANEPRCVGSGTLPTSTCTCTQATIVNVWDMSAYVKSIDPNHNVSVGDE 233
Db 189 YKDDPTIMAWELMEPRCPSD-----PSGRATQA----WITEMAAHVKSIDRHLLEAGLE 240
Qy 234 GFYIGSTQSGWPNYDPSD-GVDNNALLRVKNIDFGTVHLYPNYWGONA-----DMGTO 286
Db 241 GFY-GQSSPOSKTLNPPCOFCOTDFIANNRIPGIDFTVHVSYPDEWFPDSSSQSOMDFLNK 299
Qy 287 WIKDHANAA-AIGKPTILEEF-----GMOTPDROSVYOT-----WTQTVRTNGEAGW 333
Db 300 WLDHAIQDAONVHLKPIIAEFKSGMKKPGYTPAQRDIVFNTVYSKIYGSAKRGGAAGG 359
Qy 334 NFWMLAGNVNCPYPNY-DGNFNVYPSSTATV 364
Db 360 LFWOLL--VNG--IDNFODGYGIILSSOSSTV 387
```

Db	300	WLDHAIHQDNVLHKKPIIAEFGCKSMKKPCYTPAQORDIVFNTVYSIYGSAXKHGAAAGC	359
Qy	334	NFWMLAGNVNGQPPNY-DGFNVYPSSTATV	364
Db	360	LFWQLL-VNG--IDNFQDGYGIILSQSSSTV	387
Db	31-MAY-2002	(first entry)	
DE	Herbicide	Herbicide	
AC	Herbicide	Herbicide	
OS	Arabidopsis thaliana	Arabidopsis thaliana	
PN	WO200210210-A2	WO200210210-A2	
PD	07-FEB-2002	07-FEB-2002	
XX	28-AUG-2001	2001WO-EP09892	
XX	28-AUG-2001	2001WO-EP09892	
PA	(FARB)	BAYER AG	
PI	Tietjen K,	Weidler M	
DR	WPI:	2002-269010/31	
PT	Identifying plant target proteins for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms		
XX	Claim 5:	SEQ ID NO 3183; 261pp + Sequence Listing; English	
CC	The invention relates to identifying target proteins (ABB90790-ABB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.		
XX	Sequence	431 AA;	
Qy	Query Match	26.1%; Score 543.5; DB 23; Length 431;	
Db	Best Local Similarity	35.5%; Pred. No. 1.1e-36;	
Qy	Matches 139; Conservative 49; Mismatches 139; Indels 65; Gaps 17;		
Qy	4	GFVTASGQFVNLCLPRYGGTNNYLSY-----QSHADVDDVLAKAQAANLVSIRTWGF	58
Db	30	GFVTRGQVSLNGPYIYACFNAYILMYVASDPSQRSKISTAFQDASRHLTVARTWAF	89
Qy	59	IDIGSLDCSVPTIDGNKNGFYFYQWDPSTCAPYNDGPTGLQGLDYAIAASAAHGLRVIV	118
Db	90	SD-----GGYRALQYSPG----SYNEDM--FQGLDFALAEARRHGKIL	128
Qy	119	VLTNWDKEFGCMDOYDKW---YGLPY---HNFYTDPTQOAYKNVNHLLNRVNSITGV	173
Db	129	SFANNYESFCGRKQYVDMAKSRKPVSSDDFTDSLVLKDFYKNHIAKAVLNRFNTFTKVH	188
Qy	174	YKNDPTIFANELANPCVSGTLPTSGTCTATVNWDMQSAIYKSIDPNHMSVSGDE	233
Db	189	YKDDPTIMANELANPCPSD----PSGRAIQA----WITEMAAHVKSILDRNHLLEAGLE	240
Qy	234	GFYIGTQSGSWPNDS-DGVNNALLRVKNIDFGYHLYPNYWGQNA-----DWGTQ	286
Db	241	GFY-GOSSPOSKTLNPPGQGTDFIANNRIPGIDFVTHSYHPDEWFPDSSEQSQMDFLNK	299
Qy	287	WIKDHANAA-ATGKPTILFEF-----CWQTPDRDSVYQT-----WTQVTRTNGEAGW	333

XX PS Claim 5: SEQ ID NO 1486; 261pp + Sequence Listing; English.

XX CC The invention relates to identifying target proteins

CC (AB90790-AB94016) for herbicidally active compounds, comprising

CC aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using

CC suitable search parameters, where plant sequences having an E-value

CC greater by a factor of 3 than the E-value of most similar non-plant

CC sequences are selected. The polypeptides or nucleic acids encoding them

CC are useful for identifying modulators. The identified modulators are

CC useful as herbicides.

XX CC

XX Sequence 408 AA:

Query Match 25.8%; Score 537; DB 23; Length 408;

Best Local Similarity 33.3%; Pred. No. 3.7e-36;

Matches 135; Conservative 74; Mismatches 128; Indels 68; Gaps 16;

QY 1 APA-GEVTASGGQFVNLGPPYRGYGGTNNYLSYQ-----SHADVDVLAQAQAMNLSVIR 54

DB 24 APSDGFVRNGVQFVLNGKPPFYANGFNAYWLAYEATDPTTRFKITNVFQNAATSLGLTIAR 83

QY 55 TWGFDIGSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLGLOYATASAAHGL 114

DB 84 TWGFRD-GAIYRALQAPGS-----YDEQT-----FQGLDFVIAEAKRIGI 123

QY 115 RVIVLTLNDWKKEFGMDQYDKWYG-----LPYHDNFYDPTQOAYKNWNHLLNRVNSI 169

DB 124 KLILLVNNWDDYGGKQYVDWARSKEVSSNDDYRNPVKDFYKNHVKTVLNRVNTF 183

QY 170 TGVTYKNDPTIFAWELANPRCVGSLTPTSGTCTQATVNVWDQMSAYVKSIDPNHMY 229

DB 184 TKVAYKDEPAINAWQLMNEPRC-----GVDKSG-----KTLMDWINEMAFVKSVDPNHLLS 235

QY 230 VGDEGYIGST---QSGWPYNDPSDGVNALLRVKNIDFGTYHLYPNYWGQADMG- 285

DB 236 TGHEGYGDSSEPKNSLNPVSANTVGADFIANNHNTAIDFASMHGCSDLWFORLDQNSR 295

QY 286 -----QWIKDHIANAAT-GKPTILEFGW--QTP-----DRDSVYOT-----WTQTVRT 327

DB 296 LAFIKRWLECHIEDAQNILKPKVILAEFGILGSDTPRYTLANRDGVFTTYYDIIVYASAKG 355

QY 328 NCEAGNFWLGNVNGQYPYNDG---FNYYPSSTATVLAESA 369

DB 356 GSAAGALFW-----EVISEGMSNFAGPSSTILSDKSSSTVNIISHA 396

RESULT 12

AAG79297

ID AAG79297 standard; Protein; 416 AA.

XX AC AAG79297;

XX X: 61-JUN-2002 (first entry)

XX DE Amino acid sequence of endo-beta-mannanase enzyme of coffee.

XX Endo-beta-mannanase; coffee; polysaccharide; mannan; endo-mannanase;

XX coffee bean; transgenic plant.

XX Coffea arabica.

OS EP1138771-A1.

PN 04-OCT-2001.

PD 30-MAR-2000; 2000EP-0201175.

PF 30-MAR-2000; 2000EP-0201175.

PR (NEST) SOC PROD NESTLE SA.

XX PA

PI Marraccini P, Rogers J, Pridmore RD;

XX WPI: 2001-612542/71.

DR N-PSDB; AA165954.

XX New DNA encoding endo-mannanase of coffee, useful for hydrolysing

PT mannan polysaccharides in foods, cosmetics and pharmaceuticals ;

XX Claim 6: Page 11-13; 21pp; French.

XX The present sequence represents an endo-beta-mannanase enzyme of coffee.

CC The enzyme is involved in hydrolysis of polysaccharides that comprise

CC mannan molecules, linear or branched, linked together by beta(1-4) bonds.

CC The endo-mannanase is useful in foods, cosmetics and pharmaceuticals,

CC particularly for hydrolysis of mannan-containing polysaccharides. It is

CC also useful for synthesis and modification of such polysaccharides in

CC vitro. Treatment of coffee beans with mannanase increases the proportion

CC of dry matter extract and the extraction yield, while reducing the

CC quantity of sediment. Similar treatment of coffee liquor reduces the

CC amount of sediment produced by gelation of mannans. The endo-mannase

CC polynucleotide is useful for recombinant production of the enzyme, as

CC a source of probes and primers for in vivo detection and modification

CC of mannanase-encoding genes and for preparation of transgenic plants with

CC altered proportions of mannan-containing polysaccharides, especially

CC coffee beans, for modification of their aroma and structure.

XX CC

XX Sequence 416 AA;

Query Match 25.6%; Score 532.5; DB 22; Length 416;

Best Local Similarity 32.3%; Pred. No. 8.9e-36;

Matches 134; Conservative 65; Mismatches 127; Indels 89; Gaps 19;

QY 5 FVTASGGQFVNLGPPYRGYGGTNNYLSY-----QSHADVDVLAQAQAMNLSVIRTWGFI 59

DB 44 FVKTRTEFVNGRPLYLNGFNAYWLAYMASDPSTRTKVSTTFOQASKYQCMNAARTWAFS 103

QY 60 DIGSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLGLOYATASAAHGLRIVV 119

DB 104 DGG-----YRALQSPG---SYNEDM--FKGLDFVYSEAKYGIHLILT 142

QY 120 LTNDKKEFGMDQYDKW-----YGLPYHDNFYDPTQOAYKNWNHLLNRVNSITCVTY 174

DB 143 LVNWEYIGKKQYQVQWAKDQGHVLANDDDFDTPIVRGYFKNHIKTVLTRINSITGLAY 202

QY 175 KNDPTIFAWELANPRCVGSGTLPTSGTCTQATVNVWDQMSAYVKSIDPNHMYVSGDEG 234

DB 203 KDDPTIFAWELANPRCQSD-----LSGKAIQ-----DWISEMATHVKSIDSDHLLDGLGEC 254

QY 235 FYIGSTQSGWP-----YNDPSDGVNALLRVKNIDFGTYHLYPNYWGQADMGTO 286

DB 255 FY-----GESVPQKKEYNPGYQVGTDFISNN---RIVQVDFATIHLYPDQWPNNDQTO 306

QY 287 -----WIKDHIANAAT-AIGKPTILEEF-----QWOTPDQDS-----VYOT--WTQTVR 326

DB 307 AQFVDRWIKEHIDDSKYLLEKPLLLTEFGKSSRSPGYQAKRDAYLSHIYDTIYACAAAT 366

QY 327 TNGEAGWN-FWMLAGNVNGQYPYNY-----DGFNVYYPs--STATVLAESAIAIST 374

DB 367 GGGVCGGNLFWQVMA-----PGMESWGDGYEIVLEENPSTVGVIAQOSNRLLSS 414

RESULT 13

ABB90809

ID ABB90809 standard; Protein; 431 AA.

XX AC ABB90809;

XX DT 31-MAY-2002 (first entry)

XX DE Herbicidally active polypeptide SEQ ID NO 20.

XX KW Herbicidal; plant; agriculture; herbicide.

XX

```
OS Arabidopsis thaliana.
XX WO200210210-A2.
XX 07-FEB-2002.
XX 28-AUG-2001; 2001WO-EP09892.
XX 28-AUG-2001; 2001WO-EP09892.
XX (FARB ) BAYER AG.
XX Tietjen K, Weidler M;
XX WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms .
XX
XX Claim 5; SEQ ID NO 20; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 431 AA;
XX
XX Query Match 24.9%; Score 517.5; DB 23; Length 431;
XX Best Local Similarity 32.3%; Pred. No. 1.6e-34;
XX Matches 130; Conservative 69; Mismatches 132; Indels 71; Gaps 14;
XX
QY 4 GFVTASGQFVLNGLPYRYGNTNNYL-----SYQSHADVDVLAQAAMNLSVIRTW 56
DB 48 GFVGRNGTOFVLNCEQVYLNCFNAYWMTTAADTASGRATVTTALRQASAVGMNVARIW 107
XX
QY 57 GFDIGSLDGSVPTIDGNKNKGFYQYWDPTGAP-AYNDGPTGLOGLDYATASAAHCLR 115
DB 108 GF-----NEGDIY----PLQISPGSYSEDV--FKGLDFVYVEAGRFNIK 145
XX
QY 116 VIVVLTNDWKEFGMDQYDKWGLPYHDNFYDPTQOAYKNWVNHLLNRVNSITGVYK 175
DB 146 LIISLVNFDYGGRRKKYVEWAGLDEPDEFYNSAVQFYKNHVKTVLTRKNTITGRMYK 205
XX
QY 176 NDPTIFAWELANPRCVSGTLPSTGCTQATIV-NWVDQMSAYVKSIDPNHMYSVGDEG 234
DB 206 DDPTIFSWELINEPCNDS-----TASNILQDWKEMASVYKSDSNHLEIGLEG 256
XX
QY 235 FYICSTOGSGWPN-----DPSGVNALLRVKNTDFGTYHLYPNW-----GON 280
DB 257 FY-GESIPERTVYPPGGRVLTGDTFNN---QIPDIDFATIHYPDSWLPQSSRTGEQ 312
XX
QY 281 ADNGTOWIKDHIANA-AAIGKPTILEEF-----GNOTPDRDSVYOTWTQVTRTNGEAG 332
DB 313 DTFVDRWIGAHIEDCONIKKPLLITFEGKSSKYPGFSELEKRNFFORVYDVYDSARAG 372
XX
QY 333 WN-----FWMLAGNVNQPNVDGFNVYPSSTATVLAESA 369
DB 373 GSCTGGVFWQLTTRTGLLGDGVEFMQAGPNTTQALIAQDS 414
XX
RESULT 14
ID ABB92274
XX ABB92274 standard; Protein: 414 AA.
XX
AC ABB92274;
```

```
XX 31-MAY-2002 (first entry)
XX Herbicidally active polypeptide SEQ ID NO 1485.
XX
XX Herbicidal; plant; agriculture; herbicide.
XX Arabidopsis thaliana.
XX WO200210210-A2.
XX 07-FEB-2002.
XX 28-AUG-2001; 2001WO-EP09892.
XX 28-AUG-2001; 2001WO-EP09892.
XX (FARB ) BAYER AG.
XX Tietjen K, Weidler M;
XX WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms .
XX
XX Claim 5; SEQ ID NO 1485; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 414 AA;
XX
XX Query Match 24.8%; Score 516; DB 23; Length 414;
XX Best Local Similarity 32.3%; Pred. No. 2.1e-34;
XX Matches 133; Conservative 59; Mismatches 130; Indels 90; Gaps 14;
XX
QY 4 GFVTASGQFVLNGLPYRYGNTNNYLSYQ-----SHADVDVLAQAAMNLSVIRTWGF 58
DB 30 GFVSRKGVQFVLNGKPFYANGFNAYWLAYEATDSTTRFKITYVFQONATIHDLTIVRTWGF 89
XX
QY 59 IDIGSLDGSVPTIDGNKNKGFYQYWDPTGAPAYNDGPTGLOGLDYATASAAHCLRIV 118
DB 90 RDGYRALQI-----APGVYDEKT-FQGLDFAIAEAKRLGKIMII 128
XX
QY 119 VLTNDWKEFGMDQYDKW-----YCLPYHDNFYDPTQOAYKNWVNHLLNRVNSITGV 173
DB 129 TFVNNYSDFGGRKQYVDWAKNTQNVSSDDDFYTNPLVKQYKKNHVKTVMNRVNTFTKVE 188
XX
QY 174 YKNDPTIFAWELANPRCVSGTLPSTGCTQATIVNWVDQMSAYVKSIDPNHMYSVGDE 233
DB 189 YKDEPTIMGWELMNEPOCRAD---PSGKTLTA-----WMNEMALYVKSVDSKHLSTGLE 240
XX
QY 234 GFYIGSTOGSGWPN-----DGVNALLRVKNTDFGTYHLYPNWVGONADNGT 285
DB 241 GFY-----GSSSPORKTSLNPVAANVLGTDFIANKHKLDAIDFASISHYSPDLWFFNLDK 295
XX
QY 286 -----QWIKDHIANA-AAI-GKPTILEEFGMOT-----PDRDSVYOTWTQVTRTNGE- 330
DB 296 RLNLLRKWLECHLEDAQNILKKPLILGEFGKPTNPGVTOAQRDAVFNATFDITYESA 355
XX
QY 331 -----AGNFWMLAGNVNQPNVDGFNVYYP-----SSTATVLAESA 369
DB 356 GGPAAGALFWHVIS-----DCMNNFKDPLSVLSENSTTVNITTES 397
```

RESULT 15

AB93325
ID ABB93325 standard; Protein: 448 AA.

XX AC ABB93325;

XX DT 31-MAY-2002 (first entry)

XX DE Herbicidally active polypeptide SEQ ID NO 2536.

XX KW Herbicidal; plant; agriculture; herbicide.

XX OS Arabidopsis thaliana.

XX PN WO200210210-A2.

XX PD 07-FEB-2002.

XX PF 28-AUG-2001; 2001WO-EP09892.

XX PR 28-AUG-2001; 2001WO-EP09892.

XX PA (FARB) BAYER AG.

XX PI Tietjen K, Weidler M;

XX DR WPI: 2002-269010/31.

XX PT Identifying plant target proteins for herbicidally active compounds,
comprising aligning and comparing nucleic acid or amino acid sequences
from plant with nucleic acid or amino acid sequences from non-plant
organisms .

XX PS Claim 5; SEQ ID NO 2536; 261pp + Sequence Listing; English.

XX CC The invention relates to identifying target proteins
(ABB90790-ABB94016) for herbicidally active compounds, comprising
aligning and comparing nucleic acid or amino acid sequences from plant
with nucleic acid or amino acid sequences from non-plant organisms using
suitable search parameters, where plant sequences having an E-value
greater by a factor of 3 than the E-value of most similar non-plant
sequences are selected. The polypeptides or nucleic acids encoding them
are useful for identifying modulators. The identified modulators are
useful as herbicides.

XX SQ Sequence 448 AA;

Query Match 24.0%; Score 498; DB 23; Length 448;

Best Local Similarity 32.2%; Pred. No. 7.1e-33;

Matches 132; Conservative 66; Mismatches 130; Indels 82; Gaps 17;

QY 5 VTSGGQFVLNGLPYRGYGN:YYL-----SYQSHADVDVLAQAQAMNLSVIRTWGFD 60

QY 6 VTSGGQFVLNGLPYRGYGN:YYL-----SYQSHADVDVLAQAQAMNLSVIRTWGFD 60

QY 61 IGSIDGCVPTIDGNKNGFYFQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRVIVL 120

QY 108 -----GQHRALOKSPSVYDEEV-FRALDFVLSEARKYKIRLILSL 146

QY 121 TNDWKEFGMDQYDKW-----YGLPYHDNFTDPTQQAYKNMNVHLLNRVNSITGVYK 175

QY 147 VNNNDAYGGAQYVKGWNASCLNLTSDDDFTNPTLRNFYQSHVRTVLNRVNTFTNITYK 206

QY 176 NDTIFAWELANPRCVSGTSLTSGTCTQATIVNWVDQMSAYVKSIDPNHMYSVGDEGF 235

QY 207 NDTIFAWELMNEPRCPD-----PSGDKLQ-----SWIQEMAVEFKSLDAKHLVEIGLEGF 258

QY 236 YIGS----TQCSGWPY--NDPSGDVNNALLRVKNIDFGTYHLYPNYWGONA-----DW 283

QY 259 YGPSAPARTFNPYPAAQVGTDFIRNQVL---GIDFASVHVIPDSWISPAVSNFLEF 315

QY 284 GTQWIKDHIAAAA-IGKPTILEEF-----GWTQDPDR-----SVYOTWTOTVPTNGEA 331
Db 316 TSSWMOARVEDAEAYLGMVPLTFEGVSAHDPGFNTSPRDMHMLNTVYKMTLNSTRKGG-A 374
QY 332 GWNFMMLAGNVNGQYPNP-----YDGFNVY-----YPSSTATVLALEALAI 372
Db 375 G-----AGSLVWQVFPQGAEFMDGAYAVYLTTRAHTASKIISLQSKRLAI 418

Search completed: November 13, 2002, 11:51:39
Job time : 35.6746 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:50:41 ; Search time 17.5108 Seconds
(without alignments)
2058.756 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTASGGQFVLNGLPY.....YPPSSTATVLASEALAISTG 375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	811	39.0	669	D72278	endo-1,4-beta-mann
2	517.5	24.9	411	D86153	hypothetical prote
3	498	24.0	448	T48214	endo-1,4-beta-mann
4	495	23.8	431	T09048	probable mannan en
5	426.5	20.5	369	T04323	mannan endo-1,4-be
6	411.5	19.6	403	A84592	(1-4)-beta-mannan
7	399.5	19.2	442	T87348	mannanase, probabl
8	171.5	8.2	589	T35784	probable secreted
9	143	6.9	2817	T97033	uncharacterized pr
10	140.5	6.8	649	AH3546	outer membrane pro
11	138.5	6.7	397	A35136	cellulase (EC 3.2.
12	136.5	6.6	722	T96986	endo-1,4-beta gluc
13	134	6.4	482	JC7332	endoglycosylcerami
14	134	6.4	1684	S10789	amylase A-180 - al
15	130.5	6.3	456	C42360	cellulase (EC 3.2.
16	128	6.2	747	B47093	cellulase (EC 3.2.
17	126	6.1	493	JH0158	cellulase (EC 3.2.
18	125.5	6.0	504	S54744	cellulase (EC 3.2.
19	125	6.0	425	C97013	probable non-proce
20	124.5	6.0	2228	T97942	beta-galactosidase
21	124.5	6.0	2233	T95075	beta-galactosidase
22	123.5	5.9	1155	T71456	conserved pyrolysin
23	122.5	5.9	597	T90481	probable hypothetical
24	117.5	5.7	1144	A36968	PI-like adhesin pr
25	116	5.6	516	TJEO134	mannan endo-1,4-be
26	114	5.5	430	T55325	endo-beta-1,6-gluc
27	112.5	5.4	566	A40589	cellulase (EC 3.2.
28	112	5.4	407	T39282	probable exo-beta-
29	112	5.4	644	A43370	capsid protein - N

30	111	5.3	5188	2	B85547	probable RTX famil
31	111	5.3	5291	2	F90696	hypothetical prote
32	110.5	5.3	505	2	S39662	endoglucanase - fr
33	110.5	5.3	825	2	J50174	cellulase (EC 3.2.
34	110	5.3	563	1	CZCLBM	cellulase (EC 3.2.
35	110	5.3	1879	2	S74915	extracellular nucl
36	109.5	5.3	363	2	S30386	mannan endo-1,4-be
37	109.5	5.3	1448	2	A12007	Subtilase family p
38	109	5.2	448	2	A27631	cellulase (EC 3.2.
39	108	5.2	428	2	S03767	cellulase (EC 3.2.
40	107	5.1	409	2	B25156	cellulase (EC 3.2.
41	107	5.1	570	2	S56132	cellulase (EC 3.2.
42	107	5.1	612	2	B82756	beta-galactosidase
43	106.5	5.1	879	2	A55881	beta-mannosidase
44	106.5	5.1	1649	2	C86822	hypothetical prote
45	106.5	5.1	1742	2	T17120	cellulase (EC 3.2.

ALIGNMENTS

RESULT 1
D72278

endo-1,4-beta-mannosidase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: D72278

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: D72278

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-669 <ARN>

A:Cross-references: GB:AE001779; GB:AE000512; NID:g4981777; PIDN:AAD36302.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1227

Query Match 39.0%; Score 811; DB 2; Length 669;

Best Local Similarity 42.8%; Pred. No. 1.le-52;

Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

Qy 5 FVTASGGQFVLNGLPYRYGGTNNYLSYOSHADVDDVLAKAOAMNLSVITWGFIDIGSL 64

Db 23 FVKVENGKFPALNGKREFRFGISNNYMYHYKSNRMIDSVLESARDMGIKVLRWGF----L 77

Qy 65 DGSVPTIDGNKNGFYQWDISTCAPAYNDG-----PTGLOGLDYAIASAAHGLKIVVI 120

Db 78 DGESYCRDN-----TYMHPEPGVFGVPEGISNNAQSGFERLDYTVAKAKELGILKIVII 131

Qy 121 TNDKKEFGMDQYDKWGLPYHDNFYDTPRTOQAYKKNVHILLNRVNSTGTGYTKNDPTI 180

Db 132 VNNHDDFGGMNQYVRFVCTHDDFYRDEKIKKEYKKYVSLVNHVNTYTGVPYREEPTI 191

Qy 181 FAWELANEPKRCVSGTLPSTCTQATIVNVNDQMSAYVKSIDPNHMSVSGDEGF---YI 237

Db 192 MAWELANEPKRC-----ETDKSGN---TLVENVKENSIVKSLDNPILVAVGDEGFFSYNE 243

Qy 238 GSTGSG---WPYNDPDCVNDNALLRVKNIDFGTYHLYPNYWGON----ADWGTQWIKD 290

Db 244 GFKPYGGEAEWAYNGWS-GVDMKLLSLETVDYDCTFHLYPSHWGVSPEYNAQWAKMIED 302

Qy 291 HIAAATAIKGPTILEEFG--WQTP--DRDSVYQVOTVOTVRTNCEAGNFWMLAGNVNGOP- 346

Db 303 HIKAKEIKGPKVLEEYGIKPSAPVNTAIYRLNNDLVYDGGGAMFWMLAGICEGSDR 362

Qy 347 -----YPNYDGFNYPYPSSTATVLASE--ALAIISTG 375

Db 363 DERGYYPDYDGFIRVNDSDSPEALIREYAKLFNTG 397

RESULT 2

D86153
hypothetical protein T6A9.1 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: D86153
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
C:A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
C:Accession: D86153
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <STO>
A:Cross-references: GB:AE005172; MID:g9857528; PIDN:AAG00883.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 24.9% Score 517.5; DB 2: Length 411;
Best Local Similarity 32.3%; Pred. No. 4.5e-31;
Matches 130; Conservative 69; Mismatches 132; Indels 71; Gaps 14;
QY 4 GFTASGQFVLNGLPYRYGGTNNYLL-----SYGSHADVDVLAQAAMNLSVIRTW 56
DB 28 GFVGRNCTQFVLNGEYVYVNGFNAYMMTAAADTASGRATVTTALRQASAVGNVARIW 87
QY 57 GFIDIGSLDGSVPTIDGNKNGCFYQYWDPSGAP-AYNDGPTGLOGLDYAIASAAHCLR 115
DB 88 GF-----NEGDYI-----PLQISPCSYSEDV--FKGLDFVYVEAGRENK 125
QY 116 VIVVLTNDMKFEGGMDQYDKWGLPYHDNFYDPTQOAYKNWVHILLNRVNSITGVITYK 175
DB 126 LILSIVNWFEDYGRKKYVEMAGLDEPFTYNSAVKQFYKNHVKTVLTTRKNTITGRMYK 185
QY 176 NDPTIFAWELANPRCVGSGTLPTSGTCTQATIV-NNVDMSAYVKSIDPNHIMVSVGDEG 234
DB 186 DDPTIFSWELLNEPRCNDSE-----TASNILQDWVKEMASVKSIDSNDHILLEIGLEG 236
QY 235 FYTGSTGCGSWPYN-----DPSGVDNNALLRVKNIIDFGTYHLYPNWY-----GQN 280
DB 237 FY-GESTIPERTVYNPGGRVITGTDFITNN---QIPDIDFATIHYPDSMLPLOSSRTGEQ 292
QY 281 ADMGTQWIKDHIANA-AAIGKPTILEEF-----GWOTPDSDSVYQYQVTTQVHTNGEAG 332
DB 293 DTFVDRIWIGAHIEDCDNIHKPLLITFEFGKSSYPGHSLEKRNKFFORVYDVYDSARAG 352
QY 333 WN-----FWMLAGNVNGQPYNVDGFNFVYPSSTATVLASEA 369
DB 353 GSCTGGVFMQLTNRTGLLDGVEVFVMOAGPNTTAQIIADQS 394

RESULT 3

T48214
endo-1,4-beta-mannosidase-like protein - Arabidopsis thaliana
N:Alternate names: protein T20L15.200
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48214
R:Bevan, M.; Pecters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mey
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224488
A:Accession: T48214
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-448 <BEV>
A:Cross-references: EMBL:AL162351
A:Experimental source: cultivar Columbia; BAC clone T20L15
C:Genetics:
A:Map position: 5
A:Introns: 34/2; 126/3; 191/3; 303/2
A:Note: T20L15.200

Query Match 24.0% Score 498; DB 2: Length 448;
Best Local Similarity 32.2%; Pred. No. 1.4e-29;
Matches 132; Conservative 66; Mismatches 130; Indels 82; Gaps 17;
QY 6 VTASGQFVLNGLPYRYGGTNNYLL-----SYGSHADVDVLAQAAMNLSVIRTWGFIID 60
DB 48 VORKGMQFTLNGQPFYVNGFNMTLAADNSTRGKVKTEVPQOASAVGCMVGRWTFATND 107
QY 61 IGSIDGSVPTIDGNKNGCFYQYWDPSGAPAYNDGPTGLOGLDYAIASAAAIGLRVIVVL 120
DB 108 -----GOMRALQKSPSVYDEEV-FKALDFVLSKARKYKIKLILSL 146
QY 121 TNDMKFEGGMDQYDKW-----YGLPYHDNFYDPTQOAYKNWVHILLNRVNSITGVITYK 175
DB 147 VNMWDAYGGKAAQYVKGWNASGLNLTSDDDFTNPTLRFNYQSHVTVLNRVNTFTNITYK 206
QY 176 NDPTIFAWELANPRCVGSGTLPTSGTCTQATIVNNVDMSAYVKSIDPNHIMVSVGDEGF 235
DB 207 NDPTIFAWELMNEPRCPSD-----PSGDKLQ-----SWIQEMAVFVKSIDAKHLVLEIGEGF 258
QY 236 YIGS-----TGSGWPY---NPPSGVDNNALLRVKNIIDFGTYHLYPNWYGONA-----DW 283
DB 259 YGPSAPARTFRPNPYAAQVCTDFIRNNQVL---GIDFASVHVYPDSWISPAVSNSFLKFP 315
QY 284 GTQWIKDHIANA---ICKPTILEEF-----GWOTPDSD-----SVYQYQVTTQVHTNGEAG 331
DB 316 TSSWQMAHVEDAEYMLGMPVLTFEFGVSNHIDPGFNTSFRDMLNTVYVYKMTLNSTRKGG-A 374
QY 332 GWNEFWMLAGNVNGQPYN-----YDGFENVY---YPSSTATVLASEAALAI 372
DB 375 G-----AGSLVWQVFPQGAFFMDGQYAVYITRAITASKIISLOSRLAI 418

RESULT 4

T09048
probable mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Arabidopsis thaliana
N:Alternate names: (1-4)-beta-mannan endohydrolase; protein F26K10.200
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T09048
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Harrell, B.G.; Ban
submitted to the Protein Sequence Database, June 1999
A:Reference number: 216533
A:Accession: T09048
A:Molecule type: DNA
A:Residues: 1-431 <BEV>
A:Cross-references: EMBL:AL049803; ATSP:F26K10.200; GSPDB:GN000062
A:Experimental source: cultivar Columbia; BAC clone F26K10
C:Genetics:
A:Gene: ATSP:F26K10.200
A:Map position: 4
A:Introns: 120/3; 186/3; 227/3; 298/2
C:Keywords: glycosidase; hydrolase

Query Match 23.8% Score 495; DB 2: Length 431;
Best Local Similarity 32.9%; Pred. No. 2.2e-29;
Matches 124; Conservative 55; Mismatches 124; Indels 74; Gaps 16;

QY 1 APAGFTVTSAGQFVLNGLPYRYGGTNNYLSYQSHAD-----VDDVLAQAAMNLSVI 53
DB 37 AKLCFVKRNCTQFVDDKPLVYVGNWSYW--FMDHAYDEHSRNLVGMLEAGAKMGLTVC 94
QY 54 RTWGFIDIGSLDGSVPTIDGNKNGCFYQYWDPSGAPAYNDGPTGLOGLDYATASAAHIG 113
DB 95 RTWAF-----NDGYNALQI-----SPGRFDERV-FOALDHYVIAERKHD 133

```
Qy 114 LRVIVLTNDWKKEGGMDQYDKW-----YGL-PYHDNFYTDPTQQAYKNVNHLLNRVN 167
Db 114 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 VKLLSLVNNLQAYGKQYQVKWAWQEGVGLSSNDSFFDPISIRNYFNKYLKVLTRKN 193
Qy 168 SITGVYTKNDPTIFAWELANPRCVGSGTLPTSGTCTQATIVNVVDQMSAYVKSIDPNHM 227
Db 168 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
194 SVTGIETKNDPTIFAWELINEPRC-----TTDVSG-----KTLQDWIDENTGPIKSIDDKIL 245
Qy 228 VSVDECFYIGSTCGS-----WPNDRPSGDVNNALLRVKNIDFCTYHLYPNYNGON 280
Db 228 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
246 LTVGLEGFY-GPNSPKGLTVNPEQWASOLGTFVONS-----NSSNIDFASVHLYPDHFMH 301
Qy 281 ADWG-----TQWIKDHITANA-AAIGKPTILEEFG-----WQTPORDSVYQ-----TW 321
Db 281 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
302 QTFEKLLKVVKKWQSHLEDGLKELKPKVLPFTFGLSNONKDYEPSQDKFYRIIFDVVY 361
Qy 322 TOTVRTNCEAGNFWML 338
Db 322 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
362 KSARKRKSGAGTLVWQL 378

RESULT 5
T04323
mannan endo-1.4-beta-mannosidase (EC 3.2.1.78) - Tomato
N:Alternate names: (1-4)-beta-mannan endohydrolase
C:Species: Lycopersicon esculentum (tomato)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: T04323
R:Bewley, J.D.; Burton, R.A.; Morohashi, Y.; Fincher, G.B.
Pianta 203, 454-459, 1997
A:Title: Molecular cloning of a cDNA encoding a (1-4)-beta-mannan endohydrolase from the
A:Reference number: 215283; MUID:98083761; PMID:9421930
A:Accession: T04323
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-369 <BEW>
A:Cross-references: EMBL:AF017144; NID:q2655007; PIDN:AA887859.1; PID:q2655008
C:Function:
A:Description: participates in enzymatic depolymerization of mannose-containing polysacch
C:Keywords: glycosidase; hydrolase

Query Match 20.5%; Score 426.5; DB 2; Length 369;
Best Local Similarity 31.4%; Pred. No. 2.4e-24;
Matches 126; Conservative 52; Mismatches 126; Indels 97; Gaps 19;

Qy 1 APAGFVTASGCOQVNLGLPYRYGCTNNYLSYOSHAD-----VDDVLAKAQANLVSIR 54
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
23 ANSGFIGVKDSHFELNGSPFLNGFNSTLMIWA-ADPTERYKVTVELKDKASVAGLSVCR 81
Qy 55 TWGFIIDIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDGPTGLQGIIDYATASAAHGL 114
Db 55 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
82 TWAFSDGCG--DRALQISPGIYDERVF-----QGLDFVIAEAKKYGA 120
Qy 115 RVIVLTNDWKKEGGMDQYDKWGLPYHDNFYTDPTQQAYKNVNHLLNRVNSITGVY 174
Db 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 QI-----SND-----DEFTYHPLMKLYLKNHIEKVVTRLNSITKVA 157
Qy 175 KNDPTIFAWELANPRCVGSGTLPTSGTCTQATIVNVVDQMSAYVKSIDPNHNVSVGDEG 234
Db 175 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
158 KDDPTINAWELMNEPR-----DQADYSGKTVNGWVQEMASVFKSLDNKHLLEVMEG 209
Qy 235 FYIGSTCGSMP-----YNPDSGDVNNALLRVKNIDFGTYHLYPNYW--CONAD-- 282
Db 235 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
210 FY-----GDSIPERKSNVPCYQVGTDFISNLH---INEIDFATHATDQWVSGQSDAQ 261
Qy 283 --WCTOWIKDHIANAAI--CKPTILEEFGWQTF---PRD-----SVYQTTWTQVTRNGE-A 331
Db 283 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
262 LVNMEKWIITSHWEDARNILKPLVLAIEFGKSRQCGSRQIFMSSVYRNVYNLAKEGTM 321
Qy 332 GNFWMLAGNVNGQYPNY--DGFENVY--YPSSTATVLASEA 369
Db 332 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
322 GSLVWQLMAH-----GMENVDDGCIIVLGQTPSTTQIISDQA 358
```

RESULT 6
A84592

(1-4)-beta-mannan endohydrolase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84592
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Renito, M.L.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84592
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-403 <STO>
A:Cross-references: GB:AE002093; NID:q454480; PIDN:AAD20927.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g20680
A:Map position: 2

Query Match 19.8%; Score 411.5; DB 2; Length 403;
Best Local Similarity 29.6%; Pred. No. 3.5e-23;
Matches 113; Conservative 53; Mismatches 119; Indels 97; Gaps 16;

```
Qy 5 FVTASGCOQVNLGLPYRYGCTNNYLSYOSHAD-----VDDVLAKAQANLVSIRTWG 57
Db 42 FVKNRGTQFVVDGRKALYVNGWNSYW--FMDHAVNDHSHRVSAMLEACAKMGLTVCRTWA 99
Qy 58 FIDIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDGPTGLQGLDYAIAASAAHGLRVI 117
Db 100 F-----NDGYNALQI-----SPGRFDERV-FKALDHVIAEAKTHCVS-- 136
Qy 118 VVLTDNWKKEGGMDQYDKWGLPYHDNFYTDPTQQAYKNVNHLLNRVNSITGVYKND 177
Db 137 -----SSNDSFFDPISIRRYFNKLYLVLLTRKNSLTGLEIYND 174
Qy 178 PTIFAWELANPRCVGSGTLPTSGTCTQATIVNVVDQMSAYVKSIDPNHNVSVGDEGPI 237
Db 175 PTIFAWELINEPRCMSD-----VSGD-----TLQDWINEMTAFIKSIDNKHLLTVGLEGFG 226
Qy 238 GST-----QSGGWPYNDPSPDGVNNALLRVKNIDFGTYHLYPNYW-----GQNADWGT 285
Db 227 PSSPKKLTVPNERASELGSDFVRNS---DSPNIDFASVHLYPDHFWHDQGFEEKLKFVV 283
Qy 286 OWIKDHIANA-AAIGKPTILEEFG-----WQTPDRDSVYQTF-----WTQVTRTNGEAG 332
Db 284 KWLMSHIEDGDKELKPKVLPFTFEGFLSNLNDKYDPSORDFYRTIFDVLYKSARKKRSAG 343
Qy 333 WNFWMLAGNVNGQYPNYVDGPN 354
Db 344 TLVWQFL--IEGM-----EGFN 358
```

RESULT 7
F87348

mannanase, probable [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: F87348
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87348
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-442 <STO>
A:Cross-references: GB:AE005673; NID:g13422046; PIDN:AAK22786.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0801

Query Match 19.2%; Score 399.5; DB 2: Length 442;
Best Local Similarity 26.8%; Pred. No. 3.1e-22;
Matches 118; Conservative 65; Mismatches 159; Indels 99; Gaps 16;

QY 2 PAGFVTASGQFVNLGLPYHYGCTNNYYLSY-----QSHADVDVV---LAKAAMNLSVIR 54
DB 27 PKGFVTVKDGRLSLDGCKPKYRFAGCTNVMYAAWLACAPAGYGDGLRLRRELDRKAMGVNLR 86
QY 55 TWGFDIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDGPTGLOGLOVATIASAAHGL 114
DB 87 I-----LGAGEQSPAKV-----AMDPTFRGPCEDYNADLLKGLDVTLAEMAQRDM 131
QY 115 RVIVLTLNDMKFCGMDQYDKWYG-----LPYHDFNYDPRTOQAYKN 157
DB 132 KAVIYVNNFWDSGMPAYLNNWCDGPFQGDHPAYPHWQYADYSARFYANQKANALFRH 191
QY 158 WVNHLNLRVNSITGVYTKNDPTIFAMELANEPRKCVSGTGLTPTSGTCTQATIVNWDMSA 217
DB 192 YVTSLSVRSVSTVTKPYRDPDPTIMNSQLANEPRPGGSDAF---GQSNMPPAYQAWIRDTAG 248
QY 218 YVKSIDINHMVSGDECFYIGSTGSCWHPYNDSPDGDVNNALLRVKNIDFGTTHLYPNYW 277
DB 249 LKKRLNQHLLVSTGSG-----TWGC---MGLESCVIDAHIA---PPVIDYMTLHIWPNNW 297
QY 278 G-----GNADWGTQ---WTKDHIAAAATGKPTILEEFCW-----OTPDPR 314
DB 298 GWISMTDOPSTYAGRKCHDYVAOHIALAKRLNKLPTIEFGLIRDRGROFTPGSPPTTYR 357
QY 315 DSVYQVWTO---IVYVING--EAGNFWMLAGN-----VNGQPYPNYDG 352
DB 358 DREYKTMULDALADMKAGGPTAGVNFWMNGEGRAQQPDAPWFKKDKSYGVGDPPEQGL 417
QY 353 FNVY--YPSSTATVLAEEALAJ 372
DB 418 FGVFDADASTLAVIKAHAAV 438

RESULT 8

T35784

probable secreted protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #Text_change 05-Nov-1999

C:Accession: T35784

R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1998

A:Reference number: 221570

A:Accession: T35784

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-589 <SEE>

A:Cross-references: EMBL:AL031013; PIDN:CAA19789.1; GSPDB:GN00070; SC0E8A.17

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SC0E8A.17

Query Match 8.2%; Score 171.5; DB 2: Length 589;

Best Local Similarity 23.3%; Pred. No. 4.4e-05;

Matches 84; Conservative 38; Mismatches 108; Indels 131; Gaps 21;

QY 1 APAGFVTASGG---QFVLNGLPYRGCTNNYYLSY-QSHADVDVVLAKAAMNLSVIRT 55
DB 41 AAGSVVKTGCGGDWQTLVDGSPYTVKG-----LTWGPSADADRYLPLDASMGVNTIRT 95
QY 56 WGFIDIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDGPTGLOGLOVATIASAAHGLR 115
DB 96 W-----GTDASSRPLID-----SAAAHGVK 115
QY 116 VIVLTLNDMKFCGMDQYDKWYGLPYHDFNYDPRTOQAYKN-----WVNHLNLRVN 167
DB 116 VI---AGFWLQPGGP-----GSGGCVNYLTD---TAYKQMLAEFRPWQOE----- 156

QY 168 SITGVYTKNDPTIFAMELANE-----PRCVGSGTLPHTSGTCTQATIVNWDMSAYVKS 222
DB 157 -----YKDHGPGVLMNMGVNESVLGLQNCYGGDELEHQ-----KDAYTTFVNDVAKKIIIV 206
QY 223 DPHNMVSGDEGFYIGSTGSCWHPY---NDI-SIKVDNNALLRVKNIDFGTTHLYPNYWGO 279
DB 207 DPNHPVTSTDA--WVGA-----WVYVOKNAP-----DLDLVAV--NSYDA 242
QY 280 NADGWOWIKDHIANAATAICKPTILEEFG-----WOTP-DKDSVYQVWTOVTVRTNGEA-GW 333
DB 243 VCDVRSAAW-----EOGGYTKPYIVTSGPAGWEVPPDANGVPEEPTDRAKAOGYTDAW 296
QY 334 N 334
DB 297 N 297

RESULT 9

B97033

uncharacterized protein, related to enterotoxins of other Clostridiales [Imported]

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #Text_change 14-Sep-2001

C:Accession: B97033

R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97033

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2817 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK79053.1; PID:g15023993; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1079

Query Match 6.9%; Score 143; DB 2: Length 2817;

Best Local Similarity 22.7%; Pred. No. 0.042;

Matches 103; Conservative 57; Mismatches 162; Indels 132; Gaps 29;

QY 4 GFVTASGGQFVLN-----GLPYRYGCTNNYY-----LSYQSHADVDVIAK 44
DB 1530 GFVTLSCNTYFDSYGEIMRIGLTYI---NNNYFFNSKGIETGWSLYRYANPNGLILT 1586
QY 45 A-QAMNLSVIRTWGFDIGSLDGSVPTIDGNKNGF-----YFOYWDPTGAPAY----- 92
DB 1587 GFQTIIN---GKTYFFNSDGLSLDIQYINGSYGFQDKNGVMLYGLQITIGCTTYVLSNGI 1643
QY 93 -NDGPTGLOGLDYATIASAAAHGLRV-LVLTNDWKKEFG----- 128
DB 1644 SQSGFITLNGKTYTFDS--YVGNRTGLIONNNNYFFGDNMGTLQGWISODNLRYYANSS 1701
QY 129 -----GNDQYDKWYGLPYHDFNYDTPRTOQAYKNVNHLLNRVNSITGVYTKNDPT-IFA 182
DB 1702 GVCLTGLQITID---GKKYFFNSVARMETGLVYIN-----NTYYG--FDNDGTLIYS 1747
QY 183 WELAN-EPRCVGS-GTLPTSGTCTQATIVNWDQM--SAYVKSID---PNHMVSVGDEGF 235
DB 1748 WHNINGRMYCFNTDGTGVT-----GWINLGRSCYLDSSQGFSLTGLTLTIGNIY 1797
QY 236 YIGS---TOGSGWPNPDSGVDNNA---LLRVKNIDFGTTHLYPNYWGONADNGTOWIKD 290
DB 1798 YFGSDYSMTGWVTSGSSKYYFNESGIMLTGFTIDGNTIY-FDSY--GNSTTGTRSLN- 1853
QY 291 HIANAAAIKPTILEEFGWQTPDRDSVY---QWTQTVRTNGEAGWNFWMLAG----- 340
DB 1854 --GNCYGFNDGIMLT-GWOTISGNNYFENPDCTAKIGLNTYEGKTYYPSTGCGYTGTGII 1910
QY 341 NVNQDYP--NYDGF-----NVYPSSTATV 364
DB 1911 NINSNTYFCYDCALKTCGWIRNNYIYYADNNGII 1944


```

Qy 119 VLTNDWKEFCGMDQYDKWYGLPYHDNFPYTDPTQQAYKKWVNHLLNRVNSITCVVTYKNDP 178
      ||      : : : ||      : : : ||      : : : ||      : : : ||      : : : ||
Db 133 ----DW-----HMFENPQTD---KSKDKFFNQISS---EYKNSP 162

Qy 179 TIFAWELANPRCVCSTGLTPTSGTCTQAT----IVNVQDMSAYVKSIDPNHMVSVCDCC 234
      : : |||||
Db 163 NV1-YETANEP-----CGDVTWGRDIKPYANEVIPVIRANSINAIVIGSPT 208

Qy 235 FYIGSTGSGWPNPNDPSGDVDNNALLKVNIDFGTTHLYPNYWGQNDWGCTOMIKDHIAN 294
      : : ||||
Db 209 W-----SOSVLDPANP-----LFSNVMTAC-IFVAGTHC-----QWLRURITQ 247
      : : |||||

Qy 295 AAATGKPTILEEFCWQTPD---RDSVY---QTWTQTVRTNGFAGNFWML 338
      || : : : || : : : || : : : || : : : || : : : ||
Db 248 --ALNKKNIALFATEMGTSDCNGSGCTVIEESORWVDFM-TONKISWTNNVL 295

RESULT 13
JC7332
endoglycosylceramidase (EC 3.2.1.123) - Rhodococcus sp.
N:Alternate names: Cer glycanase; endoglycosylceramidase
C:Species: Rhodococcus sp.
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 08-Sep-2000
C:Accession: JC7332
R:Sakaguchi, K.; Okino, N.; Sueyoshi, N.; Izu, H.; Ito, M.
J. Biochem. 128, 145-152, 2000
A:Title: Cloning and expression of gene encoding a novel endoglycosylceramidase
A:Reference number: JC7332
A:Accession: JC7332
A:Molecule type: DNA
A:Residues: 1-482 <SAK>
A:Experimental source: strain C9
C:Comment: This enzyme, a member of type II enzyme, cleaves the glycosidic
C:Genetics:

```

Query Match	6.4%	Score 134;	DB 2;	Length 482;	24;
Best Local Similarity	21.1%	Pred. No. 0.021;			
Matches	87;	Conservative	59;	Mismatches 134;	Indels 132; Gaps
Qy	4	GFVTASCGQFVLNGLPYRYGCTNNYYILSYQSHADVDDYLAKAQAAANLSVIRTWGFIIDGCS	63		
Db	56	GFNTASSAKSTPDGMP-----IFTESDLREHADMGTFNFRFLTISWRSVE---	100		
Qy	64	LDGSVPTIDGNKNGFYQYWPDPSTCAPYNDGPTGLOGLDYAIASAAAHGLURVIV----	118		
Db	101	-----P-----EPGOYDQAY-----LDRVEQVRCWYALRGYKVLMDMHOD	135		
Qy	119	-----VLTDNKHFCGMDYDKWY--CLPYHDN-----FYTDPTQOAYKN-W-----	158		
Db	136	LYSGAIIPTDGGTNGCAPAWATYMDGLPVNERKDSWELYYIEPGVIRAFDQFNWTTCKHPHEL	195		
Qy	159	VNHLLNRVNSITGYTYKNDPTIFAWELANEPRCVSGTLPITSGTCQA-TIVNVWDHMSA	217		
Db	196	VDHYVNAWKAHAD-READNETVAYADLMNEP-----WGGSU--QGPAFAEAGPLTSMYQRTTD	249		
Qy	218	YVKSIDPNHNVSVCDGEFYIGSTOG-----SCHWPYNDPDCGVDDNNALLRVKNIIDFGTYHLYP	274		
Db	250	AIROVDQDSWCYVAPQA--VCVNGGIFSALCTIADPRQAGARRIAYCP-----HLYP	298		
Qy	275	-----NYWQCN-----ADMGTOWIKOHIAANAAIGKPTLIEFGWOT--PDR---	314		
Db	299	LPLDLGDGYSGFSKLTLDATIEIWTWTS--IEHVADTVLEGAPVILGERGLDITLPGAQDY	356		
Qy	315	-DSVYQYTWQTVRTNGAGCNFW-----MLAGNVNGQPYVP	348		
Db	357	LDRVY-----TVARDMGAGVYSWSDRGPMWGYLEDGCTQTITLVDTVN-KPYP	402		

RESULT 14

S10789
amylase A-180 - alkaliphilic eubacterium 163-26
C:Species: alkaliphilic eubacterium 163-26
C:Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
C:Accession: S10789
R:Cardusio, A.; Schmid, G.; Boeck, A.
Eur. J. Biochem. 191, 177-185, 1990
A:Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from a
A:Reference number: S10789; MUID:90336627; PMID:1696201
A:Accession: S10789
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1684 <CAN>
A:Cross-references: EMBL: X53373; NID:q48305; PIDN:CAA37453.1; PID:q48306

Query Match	6.48:	Score 134:	DB 2:	Length 1684:
Best Local Similarity	20.0%:	Pred. No. 0.1:		
Matches	84:	Conservative 49:	Mismatches 132:	Indels 156: Gaps 22:
OY	4	GFVTASGGQF----	VLNGLPYRGCTNNYIL----	SYOSHADVDDVLAKAQAANLSVIRK 55
DB	31	GSVEVKGQYRELNOLNENKPFSDWNNVYFVLTRDPYNGTNSDNSYGRPQ-----	IDA 84	
OY	56	WG-----	FDIGS-----	LDGSVPTIDGNKNGFYFY 82
DB	85	WGTNTIGTFHGGDKGLTRKLEEGYFTDLCTNAIWI	SAPWEQHGHW-----	GCKDCDFAHY 140
OY	83	WDPSTGAPYNDGPTGLOGLDYA-----	-----	IASAAAHGLRVI-----VLVTN--- 122
DB	141	-----	GYHGYGLDFTANDQWGTIDENREVFDLAHSGLIRVLDIDVNNHWGY 188	
OY	123	----	DWKEFGMDQYDKWGLPYLPHNFYTDPRTOQAY--	KNMVNH--LLENRVNSITGVITYK 175
DB	189	PTIVDMHEFG----	FGDTGGLP-----	RDWTPNQAQGNMIITINDINKKNEAAWANNW 238
OY	176	NDPTIFANELANEPKCVGS-----	GTLP-----	TSCTCTQATIVN--WVDOMSAYVKS1 222
DB	239	GSDWIRADETAGYDNCGGSEQTMCGFLPDI	KTEVTTGVLDLPPILRNKNWDOASGY----	294
OY	223	DPNHVSYGDEGFYIGSTQGSQWPNPDSQGVNNALRVKNI	DFTGYHLYPNYWGONAD 282	
DB	295	-----	EDWFVPAAE-----	PYRQDLNAPKDYLIK----- 319
OY	283	WCTOWIKPHIANAAAI--	CKPTILLEFCWQOTPRDRDSVYQVOTVOTVTRTNGEAGW--	NFWML 338
DB	320	WITSWEEFEGDGRVDYAKHVEIERWAEKNEAEVALQTV	RENNPNPKPGANNWDDNFWMY 379	
OY	339	A 339		
DB	380	A 380		

RESULT 15
C42360
cellulase (EC 3.2.1.4) E5 precursor - Thermomonospora fusca
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Thermomonospora fusca
C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 20-Mar-1998
C:Accession: C42360
R:Lao, G.; Ghangas, G.S.; Jung, E.D.; Wilson, D.B.
J. Bacteriol. 173, 3397-3407, 1991
A:Title: DNA sequences of three beta-1,4-endoglucanase genes from Thermomonospora fusca
A:Reference number: A42360; PMID:91258320; PMID:1904434

A:Residues: 1-466 <LAO>
A:Cross-references: GB:L01577; NID:g154693; PID:g154694
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:38-137/Domain: bacterial cellulose-binding domain homology <BCB>

```

Query Match      6.3%; Score 130.5; DB 2; Length 466;
Best Local Similarity 22.3%; Pred. No. 0.037;
Matches 82; Conservative 46; Mismatches 116; Indels 123; Gaps 21;

QY 16 NGLPYRYGCTNNVYLSYQSHADVD---DVIA---KQAMNLSVIRTWGPTDYGSLDGSVP 69
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 182 HGNPVQLRCMSTHG IQWFDHCLTDSLDALAYDKADIRLSM----- 224

QY 70 TIDGNKNGFYFQWDPSTGAPYNDGPTGL-QGLDYAIAASAAHGLRVIVVLTNDWKEFG 128
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 225 -----YIQ-----EDGYETNPRGFTDRIDQLIDMATARGLYVIV---DWHI-- 262

QY 129 GMDQYDKWGLPYHDNFYDPRTOQAYKNWVHLLNRVNSITGVTYKNDPTIFAWELANE 188
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 263 -LTPGDPHYNLDRAKTFFAEIAQRHASK-----TNVLY-----ETANE 299

QY 189 PRCVSGTLPSTGCTCTQATIVNWVDQMSAYVKSIDPNHMYVSGDEGF-YIGSTQSGWPY 247
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 300 PNCV-----SWASIKSYAEVIVPIRORPDSVLIIVGTRGSSSLGVSEGS--- 345

QY 248 NDPSCGVDDNALLRVKNIDFGTYHLYPNYWGONADGWTOMIKDHIANA---AAIGKPTIL 304
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 346 --PAEIAANP--VNASNIMY-AFHFY-----AASHRDNYLNALKEASELFPVFV 389

QY 305 EEFGMOTPDQDSV--YQTWTOTVRTNCE--AGWNFWMLAGNVNQPYPNY-DGFN---VY 356
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 390 TEFGTETYTCDCANDFOMADRYIDLMAERKIGWTKW-----NYSDDFRSGAVF 437

QY 357 YPSSTAT 363
   :| :| :|
DB 438 QPCTCAS 444
```

Search completed: November 13, 2002, 11:57:33
Job time : 19.5108 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:47:41 : Search time 8.62069 Seconds
(without alignments)
1804.220 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTAGCGFVLNGLPY.....YPSSTATVLASEALAISTG 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	179.5	8.6	562	1	GUN1_ACICE	P54583 acidothermu
2	138.5	6.7	397	1	GUN_PAEP0	P23548 paenibacill
3	128	6.2	747	1	GUND_CELFI	P50400 cellulomona
4	127	6.1	459	1	GUNA_STRLI	P27035 streptomyc
5	126	6.1	484	1	GUNA_XANCP	P19487 xanthomonas
6	125.5	6.0	504	1	GUNW_ERWCA	O59395 erwinia car
7	123.5	5.9	466	1	GUN5_THERFU	Q01786 thermomonos
8	115.5	5.6	879	1	MANB_CAPHI	Q95327 capra hircu
9	113	5.4	1953	1	BIGA_SALTY	P25927 salmonella
10	112.5	5.4	444	1	GUNN_ERWCA	O59394 erwinia car
11	112.5	5.4	566	1	GUNC_CLOTM	O05332 clostridium
12	110.5	5.3	505	1	GUNV_ERWCA	O47096 erwinia car
13	110.5	5.3	825	1	GUN3_BACSA	P19570 bacillus sp
14	110	5.3	563	1	GUNB_CLOTM	P04956 clostridium
15	109.5	5.3	383	1	MANA_STRLI	P15129 streptomyc
16	109	5.2	448	1	GUN_CLOSA	P15704 clostridium
17	107.5	5.2	426	1	GUNZ_ERWCH	P07103 erwinia chr
18	107	5.1	409	1	GUN2_BACSA	P06565 bacillus sp
19	106.5	5.1	879	1	MANB_BOVIN	Q29444 bos taurus
20	106.5	5.1	1742	1	GUNA_CALSA	P22534 caldocellum
21	106	5.1	400	1	GUN3_BACAG	O85465 bacillus ag
22	104	5.0	598	1	BGAL_XANMN	P48982 xanthomonas
23	104	5.0	603	1	BGLR_ECOLI	P05804 escherichia
24	104	5.0	900	1	GUNH_CLOTM	P16218 clostridium
25	103.5	5.0	426	1	GUN2_RALSO	P17974 ralstonia s
26	103.5	5.0	1039	1	GUNB_CALSA	P08474 c endogluc
27	102.5	4.9	651	1	BGLR_HUMAN	P08236 homo sapien
28	102.5	4.9	991	1	ENV_VILV2	P23423 visna lenti
29	102	4.9	486	1	YAIT_ECOLI	P77199 escherichia
30	102	4.9	658	1	GUN3_FIBSU	P14250 fibrobacter
31	101.5	4.9	1365	1	GTF5_STRDO	P29336 streptococc
32	99.5	4.8	499	1	GUN1_BACSU	P07983 bacillus su
33	99.5	4.8	499	1	GUN2_BACSU	P10475 bacillus su

ALIGNMENTS

RESULT 1

ID	GUN1_ACICE	STANDARD	PKT	562 AA
AC	P54583			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)			
DE	(Cellulase E1) (Endocellulase E1)			
OS	Acidothermus cellulolyticus			
OC	Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;			
OC	Actinomycetales; Frankineae; Acidothermaceae; Acidothermus			
OX	NCBI_TaxID=28049;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 43068 / 11b;			
RA	Laymon R.A., Himmel M.E., Thomas S.R.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.			
RX	MEDLINE=96346058; PubMed=8718854;			
RA	Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplun P.A.;			
RT	"Crystal structure of thermostable family 5 endocellulase E1 from			
RT	Acidothermus cellulolyticus in complex with cellobiose.;"			
RL	Biochemistry 35:10648-10660(1996).			
CC	-I- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81			
CC	DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON			
CC	CARBOXYMETHYLCELLULOSE.			
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-I- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed, usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	U33212; AAA75477.1; -			
DR	PDB; 1ECE; 14-OCT-96.			
DR	InterPro: IPR001919; Bac_celose-bind.			
DR	InterPro: IPR001547; GH_5.			
DR	Pfam: PF00150; cellulase; 1.			
DR	Pfam: PF00553; CBM_2; 1.			
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.			
FT	SIGNAL 1 41			
FT	CHAIN 42 562			
FT	DOMAIN 42 400			
FT	PRO/SER/THR-RICH (LINKER).			
FT	DOMAIN 401 461			
FT	CELLULOSE-BINDING (BY SIMILARITY).			
FT	ACT_SITE 203 203			
FT	PROTON DONOR.			
FT	ACT_SITE 323 323			
FT	NUCLEOPHILE.			

```
FT DISULFID 75 161
FT DISULFID 209 212
SQ SEQUENCE 562 AA: 60747 MW: 84E6256406A35041 CRC64;

Query Match
Best Local Similarity 8.6%; Score 179.5; DB 1; Length 562;
Matches 92; Conservative 50; Mismatches 129; Indels 113; Gaps 24;

QY 1 APAGFVTASGQFV-LNGLPYRYGCTN-----NKLVSQSHADVDVLLAKAQAAMNLSV 52
DB 42 AGCGYWHTSGREILDANNVPRVRIAGINWFGFTCNVYVHGLWSKDYKSMQDIKSLGYNT 101
QY 53 IRTWGFIDIGSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLQGLDYAIAASAAAH 112
DB 102 IRL-PTSDDIILKPGTWP-----NSINFQNMQD-----LOGLSLOVMDKIVAVAGQI 148
QY 113 GLRVIVVLTNDWKEFGMDQYDKWYGLPYHDNYTDPRTQOAYKNVNNHLLNRVNSITGV 172
DB 149 GLRII--LDRHRPDCSG--QSALWY-----TSSVSEATWISDL-----QALAQ 187
QY 173 TYKNDPTIFAMELANEPR---CVSGTLPSTGCTQATIVNW---VDMSAYVKSIDPNH 226
DB 188 RYKGNPTVCGFDLNEPHDPACWGCD-PS-----IDWRLAERACNAVLSVNPNL 237
QY 227 MVSU-----GDEFGYICSTGSG-WPYNDPSGDVNNALLRVKN-----IDFGTYHL 272
DB 238 LIFVEGVQSYNGDSYWMGNLOCAGQY-----VVLNVPNRLVYSAHDYAT-SV 285
QY 273 YPNYWGONAD-----WCTQW---IKDHANAAAGKPTILEEFGWQTPDRDSVYQTW 321
DB 286 YPOTWFSDFPNNMPCIMNKNNGYLFNQNIA-----PVLGEGF--FTLOSTTDDTW 336
QY 322 TOT-----VRTNGEAG-----WNFW 336
DB 337 LKTLVOYLRTAQYGADSTQWTFW 360

RESULT 2
GUND_PAEPO
ID GUND_PAEPO STANDARD: PRT: 397 AA.
AC P23548;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90170877; PubMed=2307659;
RA Baird S.D., Johnson D.A., Seligy V.L.;
RT "Molecular cloning, expression, and characterization of
RT endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus
RT Circulans.";
RL J. Bacteriol. 172:1576-1586(1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M33791; AAA22631.1; -
CC DR PIR: A35136; A35136.
CC DR HSRP: P54583; IECF.
CC InterPro: IPR001547; GH_5.
```

```
DR PFam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase.
FT ACT_SITE 194 194 PROTON DONOR (HY SIMILARITY).
FT ACT_SITE 317 317 NUCLEOPHILE (HY SIMILARITY).
SQ SEQUENCE 397 AA: 44357 MW: B9C3E802C04F0A2A CRC64;

Query Match
Best Local Similarity 6.7%; Score 138.5; DB 1; Length 397;
Matches 83; Conservative 48; Mismatches 143; Indels 107; Gaps 19;

QY 6 VTASGGFVNLGLPYRYGCTNNTYLSYQSHADVDVLLAKAQAAMNLSVIRKTFIDIGSLD 65
DB 47 VDESGKEAFAENGLNWFGLTETPNYTLHGLWSRSMDDMLDOVKKEGYNLIRLPYSNULFSS 106
QY 66 GSVPITDGNKNGFYQYWDPTGAPAYNDGPTGLQGLDYAIAASAAHGLRVIVVLTNWK 125
DB 107 SRPDSIDYHKN-----PDL-----VGLNPITQIMDKLIEKAGQGIQIIL----DRH 148
QY 126 EFGMDQYDKWYGLPYHDNYTDPRTQOAYKNVNNH---LLNRVNSITGVTYKNDPTIFA 182
DB 149 RPSGGGSELMYTSQYPS-----RWISDKMKLADR-----YKKNPTVIG 188
QY 183 WELANERPCVSGTLPSTGCTQATIVNW---VDMSAYVKSIDPNHMSVSGDEGFYIGS 239
DB 189 ADLNEPH--QASWGTGNAST-----DWRLAORAGNALILSVNPNMLILVEGVNINVG 241
QY 240 TQSGWPMYNDPSQVDN-NALLRVKNIDFGTYHLY-----PNYWCQ 279
DB 242 NNSQYW-WGNLTCVANYPVVLDVPRNVVYSPHIDYGPVSSQPMFNDPAPSNLPAIMDQ 300
QY 280 NADWGTOVI-KHITANAAAGKPTILEEFG-----WQTP-----DRDSVYQTW 321
DB 301 T--WG--YISKQNI-----PVLGEGFGRNVLDLSCPECKMKNALVHYIGANNLYFTY 349
QY 322 TOTVRTNGEAG-----WNFW 336
DB 350 WSLNPSNGDTGGLLDDMTTW 370

RESULT 3
GUND_CELFI
ID GUND_CELFI STANDARD: PRT: 747 AA.
AC P50400;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
DE CEND.
OS Cellulomonas fimi.
OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93209933; PubMed=8458833;
RA Weinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren K.A.;
RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RL J. Bacteriol. 175:1910-1918(1993).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- PATHWAY: Cellulose degradation.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DBL: L02544; AAA23089.1; -.
EMBL: P07986; IEXG
DR InterPro: IPR001919; Bac_celose-bind.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FNIII_repeat.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00041; fn3_2.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00553; CBM_2; 1.
DR PRINTS: PR00014; FN1YPEIII.
DR SMART: SM00060; FN3; 2.
DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 39 POTENTIAL.
FT CHAIN 40 747 ENDOGLUCANASE D.
FT DOMAIN 448 542 FIBRONECTIN TYPE-III 1.
FT DOMAIN 546 639 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 208 208 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 349 349 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 747 AA; 78936 MW; BD15473C9DB842BD CRC64;

Query Match 6.2%; Score 128; DR 1; Length 747;
Best Local Similarity 21.88; Pred. No. 0.042;
Matches 63; Conservative 41; Mismatches 119; Indels 66; Gaps 15;

Qy 94 DGPCTGLQGLDYAIAASAAHGLRVIVLTNDWKFGGMDQYDKWYGLPYHDNFYDTPRTQQ 153
Db 132 EGKNSLQIFEXWLTLCKQYGIKVFLDVHSAEADNSG-HVYNMMW-----KCDITTED 182
Qy 154 AYKNNVHLLNRVNSITGVYTKNDPTIFANELANEPKCVSGT--LPTSTCTQATIVNW 211
Db 183 VYEGW-----EWAATRWKDDDTIVGADIKNEPHGTQGTSTERAKWDGTTDKDNPKHF 233
Qy 212 VDQMSAYVKSIDPNHMYSVGDGEGYIGSTGCGWPYNDPSD--GV---DNNALLRYKNID 266
Db 234 AETASKKILAINPNWLTVFV--EGVELYKPGVPWNTSLGLTDYCTWVGGLRGVRDHPID 291
Qy 267 FGTYH-----LVPYWGQ-----NAD-WGTQWTKDHIAAAAIGKPTIL 304
Db 292 LGAHQDOLVSPHYDGPLVFDQKWFQKDFDKASLTADVWGNPLFIHDEIDIA----PLLI 347
Qy 305 EEFG---WQTPDRDSVYQTTQTVR---TNGEAGWNEFML---AGNVNG 344
Db 348 GENGCRIGQDPQDK----HWAALRDILVAERRLSOTFEWLNPNNSGDTGG 392

RESULT 4
GUNA_STRLI
ID GUNA_STRLI STANDARD; PRT; 459 AA.
AC P27035;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Endoglucanase CELA precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
DE DE
GN CELA.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
ON [1]
RC SEQUENCE FROM N.A., AND SEQUENCE OF 28-43.
RP STRAIN=66 / 1326;
RX MEDLINE=92246492; PubMed=1575483;
RA Theberge M., Lacaze P., Shareck F., Morosoli R., Kluepfel D.;
RT "Purification and characterization of an endoglucanase from
RL Streptomyces lividans 66 and DNA sequence of the gene.";
RL Appl. Environ. Microbiol. 58:815-820(1992).

```

RESULT 5
GUNA_XANCP STANDARD: PRT; 484 AA.
AC P19487;
DT 01-FEB-1991 (rel. 17, Created)
DT 15-JUN-2002 (rel. 41, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Major extracellular endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-
DE beta-glucanase) (Cellulase).
GN ENGXA OR XCC3521
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-61.
RX MEDLINE=90323605; PubMed=2373365;
RA Cough C.L., Dow J.M., Keen J., Henrissat B., Daniels M.J.;
RT "Nucleotide sequence of the engXA gene encoding the major
RL endoglucanase of Xanthomonas campestris pv. campestris";
RL Gene 89:53-59(1990).
RN [2]
RP SEQUENCE FROM N.A.
KC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorfy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsighieri A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madelira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
KL Nature 417:459-463(2002).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION OF THE PROTEIN IS NOT CRUCIAL
CC FOR ACTIVITY.
CC -1- MISCELLANEOUS: THE THR/PRO-RICH REGION (ALSO TERMED "HINGE") MAY
CC BE A POTENTIAL SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M32700; AAA27612.1;
CC DR EMBL: AE012472; BAM42791.1;
CC DR PIR: JH0158; JH0158.
CC DR HSSP: P54583; LECE.
CC DR InterPro: IPR001547; GIL5.
CC DR Pfam: PF00150; cellulase.1.
CC DR PROSITE: PS00659; GLYCOSYL-HYDROL-F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 484 MAJOR EXTRACELLULAR ENDOGLUCANASE.
FT DOMAIN 375 399 THR-PRO REPEATS ("HINGE") (PRO-THR BOX).
FT ACT_SITE 182 182 PHOTON DONOR (BY SIMILARITY).
FT ACT_SITE 303 303 NUCLEOPHILE (BY SIMILARITY).
FT -----

FT CONFLICT 410 484 VDSNMGYCNRVQVNTGTASCTWSIAVIVTCTVNNAMNA
FT TWSQSCSTLRASCVDENRTLAACATAFEFGCAAS -> ASP
FT VVGSAAARKLPAASRLACHWPPASSTGHRVWVIAAPSVTWKRP
FT HAARAIERRMVRVTKLRATRLNRKGPLPAPAIITTOHAPRL
FT (IN REF. 1).
SO SEQUENCE 484 AA: 52241 MW: 6671AE5BF1B7602A CRC64:
Query Match 6.1%; Score 126; DB 1; Length 484;
Best Local Similarity 21.9%; Pred. NO. 0.035;
Matches 92; Conservative 133; Indels 144; Gaps 26;
QY 4 GFVTASGGQVFLNGLPYRGCTNNYLSYQSHADVDVLAQAQANLSVIRKMGFIDIGS 63
DB 51 GFET---GNVHMGLMAR-----NWKDMIVOMOGGLGFNAVRL-PFCPATL 91
QY 64 LDGSVP-TIDGNKNGFYQWDPSTGAPAYNDGPTGLOGLDYATASAAHGLRIVIVLTN 122
DB 92 RSDTMPASIDYSRN-----ADLQGLTSLQILDVKVIAEFNAKGM--YVLLDH 135
QY 123 DWKEFGGMDQYDKWGLPYHDNFYTDPRTOQAYKNVNNHLLNRVNSITGVTKNDPTIFA 182
DB 136 HTPDCAGISEL--WY-----TGSYTEAQ-----WUADLRFVAN-----RYKNVPVLG 176
QY 183 WELANPRCVGSGTLPTSGTCTQATIVNW---VDMSAYVKSIDPNHMSVGDGFPY--- 236
DB 177 LDKLNEPH--GAATWGTGNAAT-----DNKAAERGSAALAVAPKWLJAV--EGITDNP 227
QY 237 IGSTQGS-GW-----PYNDPSGCVNNALLRVKNIDFGTYHILYPNWGON----- 280
DB 228 VCSNNGIGFVGGLNLOPLACTPLNIPA-----NRLLAPHV-YGPDVEVQSYFNSNFPNN 281
QY 281 --ADWGTQMTKDHIAAAALCKPTILFEFGWQTPDRDSVYQTWTQTV----- 325
DB 282 MPALW-----ERHGFQFAGT-HALLLGEFGKYGEGDARDKTQDALVKYLRKSGINUGF 335
QY 326 -----RTNGEAG-----W-----NFMLAGNVNGQPYPNYDGFNVVYPSSTA 362
DB 336 YWSWNPNSGDTGGILRDDMTSVRODKMTLLRLTGWAGTAGNTTPTPTPT-----PIPTP 387
QY 363 T 363
DB 388 T 388
RESULT 6
GUNW_ERMCA STANDARD: PRT; 504 AA.
AC Q59395;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase; V1)
DE (Cellulase V1).
GN CELV1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinhelmo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celV1 and the role of cellulase in
RL phytopathogenicity";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----


```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X79241; CAA55823.1; -
CC HSSP: O85465; IAH3.
CC InterPro: IPR001956; CHD_3.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF00942; CBM_3; 1.
CC ProDom: PD001947; CHD_3; 1.
CC ProSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation: Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 504 ENDOGLUCANASE V1.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 188 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 6.0%; Score 125.5; DB 1; Length 504;
Best Local Similarity 24.6%; Pred. No. 0.04;
Matches 62; Conservative 23; Mismatches 76; Indels 91; Gaps 16;

Qy 107 ASAAHGLRVIVLTNDWKFGMDQYDKWYGLPYHDFNYTDPRTQOAYKNWVHLLNRV 166
Db 114 AVAAAGLGVVILI--DWHTLS-----DN---DPNT---YKAQAKIFFAEM 151
Qy 167 NSITGVYTKNDPTIFANELANEPKVCSTGLPTSGTCTQATIVNWDQMSAY-----V 219
Db 152 AGL-----YGNSPNVI-YEIANEPN--GS-----VTMNGQIRPYALEVTDTI 190
Qy 220 KSIDPNHMVSGDEGFGYIGSTQSGWPNYNDPSDGVN-----NALLRVKNIDFGCTYHLP 274
Db 191 RSKDPDNLIIV-----CGSTWSQDIHDAADNQLPDPNTLYAL-----HFYA 231
Qy 275 NYWGONADMTQWIKDHIANAAIGKPTILEEFGWQTPDRDS-----VYQTWTQTVRT 327
Db 232 GTHC-----QFLRDIRYQAQRGAALFVSE--WGTSDASGNGCPFLPESQTWIDFLNN 282
Qy 328 NGEAGNFWMLA 339
Db 283 RG-LSWYNWSLS 293

RESULT 7
ID GUN5_THIEFU STANDARD; PRT; 466 AA.
AC Q01786;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase E-5 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-4)
DE (Cellulase E-5) (Cellulase E5).
GN CELE.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=91258320; PubMed=1904434;
RA Lao G., Chantas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RT Thermomonospora fusca."
RL J. Bacteriol. 173:3397-3407(1991).

```

```

RN [2]
RA REVISIONS.
RA Lao G., Chantas G.S., Jung E.D., Wilson D.B.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 137-142 AND 157-166.
RC STRAIN=YX;
RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
RT "Activity studies of eight purified cellulases: specificity,
RT synergism, and binding domain effects."
RL Biotechnol. Bioeng. 42:1002-1013(1993).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L01577; AAC09379.1; -
CC PIR: C42360; C42360.
CC HSSP: P07986; IEXG.
CC InterPro: IPR001919; Bac_celose-bind.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF00553; CBM_2; 1.
CC ProSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation: Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 36
FT CHAIN 37 466 ENDOGLUCANASE E-5.
FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 466 AA; 49800 MW; 1CF0ADFBF2DEF82E CRC64;

Query Match 5.9%; Score 123.5; DB 1; Length 466;
Best Local Similarity 22.1%; Pred. No. 0.052;
Matches 81; Conservative 46; Mismatches 117; Indels 123; Gaps 21;

Qy 16 NGLPYRGCTNNYVLSYQSHADV---DVLA---KAQAMNLSVIRTWGFIDIGSLDGSVP 69
Db 182 HGNPVQLRGWSTHGQWFDHCLTDSSLDALAYDKWAKDITRLSM----- 224
Qy 70 TIDGNKGFYFOYWDPTSGAPAYNDGPTGL--OGLDYATASAAHGLRVIVVLTNDWKEFG 128
Db 225 -----YIQ-----EDGYETNPRGFTDRHHQLIDMATARGLYVIV---DWHI-- 262
Qy 129 GMDQYDKWYGLPYHDFNYTDPRTQOAYKNWVHLLNRVNSITGYTKNDPTIFAMELANE 188
Db 263 -LTPGDPIYNLDRAKTFEFAEIAQRHASK-----TNVLY-----ETANE 299
Qy 189 PRCVSGSLTPTSGTCTQATIVNWDQMSAYVKSIDPNHMVSGDEGP-YIGSTQSGWPY 247
Db 300 PNGV-----SWASTKSYAEVPIVIRQRPDPSVIVITGRWSSLCGVSEGC--- 345
Qy 248 NPSDGVDDNNALLRVKNIDFGCTYHLYPNYWGONADMTQWIKDHIANAAIGKPTIL 304
Db 346 --PAEIAANP--VNASINMY-AHFY-----AASHRDNYLNLARASELFPVVFV 389
Qy 305 EFGWQTPDRDSV--YQWTQTVRTNGE--AGMFWMLACHNVNQCYPNY-DGFN---VY 356
Db 390 TEFCTYTYTGDNDFQMDRYIDLMAERKIGWTKW-----NYSDDFRSGAVF 437
Qy 357 YPSSTAT 363
Db 438 QPGTCAS 444

```

RESULT 8
MANR_CAPHI STANDARD; PRT: 879 AA.
AC Q95327;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE beta-mannosidase precursor (EC 3.2.1.25) (Mannanase) (Mannase).
GN MANBA.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97079658; PubMed=8921369;
RA Leiprandt J.R., Kraemer S.A., Halthcock B.E., Chen H., Dyme J.L.,
RA Cavanagh K.T., Friderici K.H., Jones M.Z.;
RT "Caprine beta-mannosidase: sequencing and characterization of the
RT cDNA and identification of the molecular defect of caprine
RT beta-mannosidosis";
RL Genomics 37:51-56(1996).
CC -|- FUNCTION: EXOGYCOSIDASE THAT CLEAVES THE SINGLE BETA-LINKED
CC MANNOSE RESIDUE FROM THE NON-REDUCING END OF ALL N-LINKED
CC GLYCOPROTEIN OLIGOSACCHARIDES.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC mannose residues in beta-D-mannosides.
CC -|- PATHWAY: PENULTIMATE STEP IN N-LINKED OLIGOSACCHARIDE CATABOLISM.
CC -|- SUBCELLULAR LOCATION: Lysosomal.
CC -|- TISSUE SPECIFICITY: FOUND IN SPLEEN AND TO A LESSER EXTENT IN
CC LIVER. NOT DETECTED IN KIDNEY OR BRAIN.
CC -|- DISORDER: DEFECTS IN MANBA CAUSE BETA-MANNOSIDOSIS, A SEVERE
CC RESULTING IN TREMOR, NYSTAGMUS, ATAXIA AND EARLY DEATH. THE
CC PRIMARY STORAGE PRODUCTS ASSOCIATED WITH THE ENZYME DEFICIENCY
CC ARE THE TRISACCHARIDE MAN-BETA-1-4-GLCNAC-BETA-1-4-GLCNAC AND THE
CC DISACCHARIDE MAN-BETA-1-4-GLCNAC.
CC -|- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U46067; AAC48665.1;
DR InterPro: IPR001649; GH_2.
DR Pfam: PF02836; Glyco_hydro_2_C; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal; Polymorphism.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 879 BETA-MANNOSIDASE.
FT ACT_SITE 457 457 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 56 56 R -> Q.
FT VARIANT 340 340 T -> S.
SQ SEQUENCE 879 AA; 101385 MW; 2896F03596B480C CRC64;
Query Match 5.6%; Score 115.5; DB 1; Length 879;
Best Local Similarity 20.4%; Pred. No. 0.43;
Matches 62; Conservative 44; Mismatches 85; Indels 113; Gaps 17;
QY 78 FYFQY-----WDSTCAPYNDGPTG--LOGLDYATASAAHGLRVIVLTNDW 124
DB 341 FYFKINGLPIFLKGNWIP--ADSFQDRVTSMDLRLLQLQSVVDAMNNAIRV-----W 390
QY 125 KEFGGDQYDKWY-----GLPYHDNFY-----TDPRTQAYKNWVHLLNRVNSITG 171

Db 391 G--GGIYEODEFELCDLGLIMIWODPWFACALYPTDEDFMDSVREEVTHQVRKL----- 443
QY 172 VTYKNDPTIFAWELANPRVCVSCSLPTSGTCTOATIVNWV-----DUMSAYV 219
Db 444 ---KSHPSIITWSGNNEEA-----ALMMGWYDTKPGYLITVIKDYVTLV 486
QY 220 KSIDPNHIVSVGDE--GFYIGS-----TOGSGHPYNDPSCVDNNALLRVKNIDFCTYH 271
Db 487 KNI--RTIVLEGDDQTRPFIISSTPNGAKTTAEGWLSNPYD-----LNYGDVH 512
QY 272 LYPNYWQONADWGTQWIKDHFANAAIGKPTILFEFGWQTPDRDSVYQVMTOTVKTNGEA 331
Db 533 FY-DYMSDCNWRV-----FPKARFVSEYGYQS-----WPSFSTLEKVSSE 573
QY 332 GWNF 335
Db 574 DWSY 577
RESULT 9
BIGA_SALTY STANDARD; PRT: 1953 AA.
AC P25927; P25928; Q9XCQ3;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative surface-exposed virulence protein bigA precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Salmonella
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 14028;
RA Stojiljkovic I., Valentine P., Heffron F.;
RT "Salmonella typhimurium rhs homolog";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-765 FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=911100301; PubMed=19871123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cysG plasmid to overcome limiting shiroheme
RT cofactor";
RL J. Bacteriol. 173:325-333(1991).
CC -|- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 414 and 732.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF133696; AAD39458.1;
DR EMBL: AE008859; AAL22340.1;

DR EMBL: M64606; AAA27042.1; ALT_FRAME.
DR EMBL: M64606; AAA27043.1; ALT_FRAME.
DR PIR: C39200; C39200.
DR PIR: D39200; D39200.
DR StyGene: SG10437; bigA.
KW Virulence: Repeat: Signal: Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 1953 PUTATIVE SURFACE-EXPOSED VIRULENCE
FT PROTEIN BIGA.
FT DOMAIN 101 252 15 x 11 AA TANDEM REPEATS.
FT REPEAT 101 103 1 (INCOMPLETE).
FT REPEAT 104 113 2 (INCOMPLETE).
FT REPEAT 111 122 3 (INCOMPLETE).
FT REPEAT 123 133 4.
FT REPEAT 134 144 5.
FT REPEAT 145 155 6.
FT REPEAT 156 166 7.
FT REPEAT 167 177 8.
FT REPEAT 178 188 9.
FT REPEAT 189 199 10.
FT REPEAT 200 210 11.
FT REPEAT 211 221 12.
FT REPEAT 222 232 13.
FT REPEAT 233 243 14.
FT REPEAT 244 252 15 (INCOMPLETE).
FT CONFLICT 207 207 D -> DRGDDDDTPDD (IN REF. 1).
FT CONFLICT 514 514 A -> R (IN REF. 3).
FT CONFLICT 1698 1698 D -> N (IN REF. 1).
FT CONFLICT 1795 1798 QYLE -> ITLQ (IN REF. 1).
FT CONFLICT 1836 1837 SA -> T (IN REF. 1).
SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match 5.4%; Score 113; DB 1; Length 1953;
Best Local Similarity 20.5%; Pred. No. 1.7; Indels 184; Gaps 23;
Matches 106; Conservative 47; Mismatches 179; Indels 184; Gaps 23;

QY 9 SGGQFVL--NGLPYRYGCTNNYLSYQSH-----ADVDDVL-----42
DB 908 AGGDTLLISDQPGSGGIEVYPLKWTHTFYAMMASDYGVVDNDEGATIHLOGAGVYGV 967
QY 43 --AKAAMNLSVIRTWGFDIGSLDGSVPTIDGNKNKGYFYQWDPST-----87
DB 968 TASRKALNEGNI-----YLDGLVPTLDDENNIITSTSYWPPSSLYLTSNGMVGST 1018
QY 88 -----GAPAYNDGPTGLGLOL-----DVAIASAAAHG 113
DB 1019 DAGDATAINTGNITVNNAGFGMALNGCTAINCGVITLTADDGVTGQADELVGMAALNG 1078
QY 114 LRV!-----VVLTDNMFEG--MDQYDKWYGLPYHDNPFY-- 146
DB 1079 GVVINDTSGVINIDADYGOAFSLDSSSVYIINNGSINLNGSPMDDTDSHMGCTPTDKIWIQ 1138
QY 147 -----TDPRTOQ-----AYKNVNHLLNRVNSITGVYKNDPTIFANELANEPKCV 192
DB 1139 SLPGSGDSDDRTSOTGPTTAGTILANYGTETLNGDDVDNG-----GW-LYNEAGAS 1187
QY 193 -----GSGTLPTSGTCTQATIVNWDQMSAYVKSIDPNHMSVSGDEGFY-----236
DB 1188 LTVNGTIVTNGANALANYGLDADAISTHSLFNEADGSIITDILLTNGDVTFFNNGDF 1247
QY 237 IGS7QSGGWPYNDSDGVNALLRVKNIDFGTYHLYPNYWGQONADGWTQWKDHANAA 296
DB 1248 TGS1AGTSYQOEIVNTG--DMTVAEDGKSLVSGSFY-----NEEDAT---LTNMSGAV 1297
QY 297 AIGKPTILEEFGWTPDRDSVYQWOTQTV-RTNG-----EAGWNFWMLAGVNG 344
DB 1298 EGGENTIN-----LTRANSLTQVNSGTITATNGYSATITVNGSNDPKWIWNTATGVING 1353
QY 345 ----QYPNYD-GFNYYVPSSTATVLASEALAI 375
DB 1354 INPDAPLINLGRGVN-FGNQCTINVQGDNAVAISGG 1388

RESULT 10
GUNN_ERWCA STANDARD: PKT: 444 AA.
ID Q59394;
AC Q59394;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (kndo-1,4-beta-glucanase N)
DE (Cellulase N).
DE GN CELN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATROSEPLICA FCBR C18;
RX MEDLINE=98299944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duns J.O., Svendsen L., Weygner C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
one multi-enzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L39788; AAC37033.1; --
CC HSP; 085465; IAS3.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase: 1.
CC Pfam: PF00942; CBM_3: 1.
CC PRODOM: PD001947; CBD_3: 1.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5: 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 444 ENDOGLUCANASE N.
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 444 AA; 48300 MW; F47E4179004CBB43 CRC64;

Query Match 5.4%; Score 112.5; DB 1; Length 444;
Best Local Similarity 22.4%; Pred. No. 0.32;
Matches 61; Conservative 27; Mismatches 89; Indels 95; Gaps 15;

QY 87 TCAPAYNDGPTGLGLOLDYAIASAAAHGLRVIVLTDNMFEGMDOYDKWYGLPYHDNPFY 146
DB 96 TAADGYISNPLANKVKEAFAAASGLGVYIIII---DMHILS-----DN---135
QY 147 TDPRTOQAY-KNVNHNLLNRVNSITGVYKNDPTIFANELANEPKCVSGTLPTSGTCTQ 205
DB 136 -DPNYYKAQKATFFAEMAGLYGSSPNVIY-----ETANE-----PNGG----172
QY 206 ATIVNWDQMSAY-----VKSIDPNHMSVSGDEGFYIGTQSGGWPYNDPSDGVN---256
DB 173 ---VTNNGQIRPYALEVTDITRSKDPDNLIIV-----GTGTWSODLHDAADNQL 218
QY 257 ---NALLRVKNIDFGTYHLYPNYWGQONADGWTQWKDHANAAAIKPTILEEFGWTPD 313
DB 219 PDPNTLYAL-----HFYAGTHG-----QFLRDRIDYAQSRGAIFVSE--WGTSD 261
QY 314 RDS-----VYQWTQTVRTNGEAGWNFWML 338

Db 83 DVKTILTSIADRGINILRIPISTELLISW---MIGK-PNPVSSVTSANNUPYHVWNUDHY 144

Qy 84 DPSTGAPAYNDGPTGLQGLDYATIASAAHIGLRVIVLITNDWKKEFGMDQTDKWKYGLPYIID 143

Db 139 DPET-----DDVKNSMEIFDIIMCYCKELGIKVMIDIHSPDANNSG-HNVELMYGKE-- 189

Qy 144 NFYTDPRTOOAYKNNWHLNKRNSITCVYTKNUPTTIFAMELANEPKCVSGTILPTSGTC 203

Db 190 ---TSTCGVVTWKMIIDTLWLADK-----YKNDDTIIAFDLKKEPH----- 228

Qy 204 TQATIVNKKVQMSAYVKSIDPNHIVMSVSGEGF-----YIGSTGSGWPYNDPSDG 253

Db 229 -----GKRGYIAEVPKLLAKWDNSTDENNKVAAETCA 261

Qy 254 -----VONNALLRVKNI-----DFGTYILYPNVWGQAD---WGQW-----IKD-- 290

Db 262 KATILEVNPKVLIVTEGVQPKTEKGYTYDTPDWGATGDASPMYSAMWGNLRGVKDYD 321

Qy 291 -----RIANAAIGKPTILEEFGWQTPDRDSVQWOTM 321

Db 322 IDLGPLNSQIVYSPHDYGPSVYAPWPEKQFTMOTLLDDYWDYTM 366

RESULT 12

ID	GUNV_ERWCA	STANDARD;	PKT;	505 AA.
AC	Q47096;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-D-glucanase V) (Cellulase V).			
GN	CELV.			
OS	Erwinia carotovora.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Pectobacterium.			
ON	NCBI_TaxID=554;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SCRI93;			
RX	MEDLINE=94067016; PubMed=8246888;			
RT	Cooper V.J.C., Salmon G.P.C.;			
RT	"Molecular analysis of the major cellulase (CelV) of Erwinia			
RT	carotovora: evidence for an evolutionary 'mix-and-match' of enzyme			
RT	domains";			
RL	Mol. Gen. Genet. 241:341-350(1993).			
CC	-1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE pH			
CC	OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES			
CC	CELSTUS.			
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X76000; CAA53592.1; -			
DR	HSSP; O85465; 1A3H.			
DR	InterPro; IPR001956; CBD_3.			
DR	InterPro; IPR001547; GH_5.			
DR	Pfam; PF00150; cellulase; 1.			
DR	Pfam; PF00942; CBM_3; 1.			
DR	ProDom; PD001947; CBD_3; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal.			
KT	SIGNAL.			
KT	1			
KT	31			

DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.					
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.					
FT	SIGNAL	1	28	POTENTIAL.		
FT	CHAIN	29	825	ENDOGLUCANASE C.		
FT	ACT_SITE	219	219	PHOTON DONOR (BY SIMILARITY).		
FT	ACT_SITE	335	335	NUCLEOPHILE (BY SIMILARITY).		
SO	SEQUENCE 825 AA; 92015 MW; A1727DA3D7632617 CRC64;					
Query Match 5.3%; Score 110.5; DB 1; Length 825;						
Best Local Similarity 19.6%; Pred.No 0.94;						
Matches 79; Conservative 44; Mismatches 122; Indels 159; Gaps						
Qy	-	57	GFTDIGSLGDSVPTIDGNKNGFYFQYWDSPSTGCAPAYNDG-PTGCLGDLGYAIAASAAHIGLR I15			
Dd	I :	:	I I I	I I I : :	I I I:	I I I:
Dd	74	GALQLVEVDGVTLAD-----ODGVPIQLRCM-----STHGLQ I06				
Qy	I16	-----VIVVLTDNW-----KEFGCMD---OYDKKWYGI, I39				
Dd	I07	WFGEIENVENAFALANDSGSNVRLALYICENAYRYNPOLIEKVYAGIELAKENDHWYII I66				
Qy	I40	PYHDNFYTPRTOAQYKNWN-----HLNRVNISITCVTVKYNDDPTIFAWEL I85				
Dd	I67	DWHVHADCDPNA-DIYOGGVNDECEYLCAKDFFLIAGE-----YPNDPHLII-YEL I216				
Qy	I86	ANEPRVCVGS-----TLPTSCTGTQTATIV----NKVDQM--S I216				
Dd	I217	ANEPSNSSGGPGITNDEDGWAEVYAQPILDALRDSCNAEDNIIVGSPNNWSQHMDLA I276				
Qy	I217	AYVKSIDPNHMVSVDGEGFYIGTQGSGWPYNDPSDCVD-NNALRVK-NIDFGTYHI,YP I274				
Dd	I277	AADNPIDHHI--TWYTLHFYTGTHEGTSYESPEGISSEDRSNVMAKYVALDKGK-AIPA I333				
Qy	I275	NYMG-----ONADMGTOMIKDHITANAAILCKPTILEEFCWO I310				
Dd	I334	TENGVSSEADGNGGPYLEADVWLFLNENNISM-TWNSLTNNKETSGAFTPILNESDAT I392				
Qy	I311	--TPDRSDVVYQWT-QTVRTNGEAGFNFWMLAGNVACOPYPYND I351				
Dd	I393	DLDPGED---QVNSMEELSVSQE-----YVKSRIJLEEYQPID I427				
RESULT 14						
ID	GUNB_CLOTM	STANDARD;	PRT;	563	AA.	
AC	P04956;					
DT	13-AUG-1987 (Rel. 05, Created)					
DT	13-AUG-1987 (Rel. 05, Last sequence update)					
DE	01-FEB-1995 (Rel. 31, Last annotation update)					
DE	Endoglucanase B precursor (EC 3.2.1.4) (ECB)					
DN	(Cellulase B).					
GN	CBLB,					
OS	Clostridium thermocellum.					
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;					
CC	Clostridium.					
RX	NCBI_TaxId=1515;					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=NCIB 10862;					
RX	MEDLINE=86148508; PubMed=3453102;					
RA	Grepinet O., Beguin P.;					
RT	"Sequence of the cellulase gene of Clostridium thermocellum coding					
RL	for endoglucanase B.;"					
RT	Nucleic Acids Res. 14:1791-1799(1986).					
CC	-I- FUNCTION: THIS ENZYME CATALYZES THE ENDOTHERMOHYLYSIS OF 1,4-BETA-					
CC	GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-					
CC	GLUCANS.					
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic					
CC	linkages in cellulose.					
CC	-I- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS					
CC	WELL AS IN OTHER C.THERMOCELLUM CELLULOSOME ENZYMES. THIS DOMAIN					
CC	MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.					
CC	-I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL					
CC	HYDROLASES). ;					

```
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03592; CAA27266.1; -.
DR PIR: A23512; C2CLBW.
DR HSP: P54583; ICEE.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00404; Dockerin_1; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00448; CLOS_CELLOUSOME_RPT; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 563
FT ACT_SITE 204 204
FT ACT_SITE 363 363
FT DOMAIN 502 557
FT REPEAT 502 526
FT REPEAT 534 557
FT REPEAT 534 557
SQ SEQUENCE 563 AA; 63929 MW; 866FE55704A1DE4B CRC64;

Query Match 5.3%; Score 110; DB 1; Length 563;
Best Local Similarity 21.8%; Pred. No. 0.65;
Matches 52; Conservative 35; Mismatches 81; Indels 70; Gaps 14;

OY 83 WDPSTGAPAYNGPTGLOG---LDYATASAAAGHLRVITVLTNDKKEFGMDQYDKWYGL 139
DB 114 YPSTDTSYNNPALAGLSYELFMENFRVGIKVLIDVHSPETDNGQ-HNTPPLMTNT 172

OY 140 PYHONEYDTPRTOQAYKN---WVHLLNRVNSITGVTYKNDPTIFAWELANEPRCVSGT 196
DB 173 TI-----TEELFKAWVW-----AERYKNDTIIIGDLKNEPH-TNIGT 211

OY 197 LPTSGTCTQATII-----VWV---VDOMSAYVKSIDPNHMSVGDGEFYIGTQGSWP 246
DB 212 MKIK---AQSAIWDSDSNHNPKNVKAETALAILVHPNVLIFVEGVEMV----- 258

OY 247 YNDPSGDVNNALLRVK---NIDFGTYHLYPNYWGONADMGCTQWIKDHANAAAAIGK 300
DB 259 ---PKDGIWDETDTSPWTCNDY-----YGNWNGN-----LRCKVLDYPIN---LGR 301

RESULT 15
MANA_STRLI STANDARD; PRT; 383 AA.
AC P51529;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (Beta-
DE mannanase) (1,4-beta-D-mannan mannanohydrolase).
GN MANA.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomyces; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 36-42.
RC STRAIN=66 / 1326;
RX MEDLINE=93207541; PubMed=8457214;
RA Arcand N., Kluepfel D., Paradis F.W., Morosoli R., Shareck F.;
RT "Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence
RT of the manA gene and characterization of the enzyme.";
RT Biochem. J. 290:857-863(1993).
```

```
[2]
RN RP REVISIONS TO C-TERMINUS.
RC STRAIN=66 / 1326;
RA Shareck F.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES
CC CELSIUS AND A PH OF 6.8.
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannans, galactomannans, glucomannans, and
CC galactoglucomannans.
CC -!- SUBUNIT: MONOMER.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M92297; AAA26710.2; -.
DR InterPro: IPR002883; CHD_5.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF02013; CBM_10; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 35
FT CHAIN 36 383
FT DOMAIN 336 340
FT DOMAIN 336 340
SQ SEQUENCE 383 AA; 39681 MW; 5DB4B407C64E94C3 CRC64;

Query Match 5.3%; Score 109.5; DB 1; Length 383;
Best Local Similarity 20.5%; Pred. No. 0.44;
Matches 90; Conservative 54; Mismatches 113; Indels 181; Gaps 27;

OY 1 APAGFTVAGSGQFVLNGLPYRGYGTNNYLSVQSHADVDVLLAKAQAAMNLSVIRTWGTFD 60
DB 4 ARSLTITTAGMAFAVLGLLFPALAGPS-----AGRAEA-----AAGGIIH 41

OY 61 IGSLOGSVPTIDGNKNGFYFO-----YMDPSTGAPAYNDGPTGIGLDIYAIASAAHGL 114
DB 42 VS--NGRV--VEGNGSAFMRGVNHIAITWYTPDR-----TG-----STADIAAKGA 82

OY 115 RVI-VVLTNDKHEFGMDQYDKWYGLPYHDNFYTPRTOQAYKNVHLLNRVNSITGV 173
DB 83 NTVRVVLJSS-----GG-----RWTKYSASEVSALICOC 110

OY 174 YKNDPTIFAWELANEPRCV-----GSGTLPTSGTCTQATIVNWVDMSAYVKS 221
DB 111 -----KANKVICVLEVHDTTGYKDGATSLDQGD-----YWGVKSAAMRA 152

OY 222 IDPNHMSVGDGEFGYIGTQSGWPYNDPSGVNUNALLRVKNIDFGTYHLY--PNYWGQ 279
DB 153 QEDYVVVNIQNEPF--GNTNYAAM-----TDATKSAICKLRGAGLHMYDAPN-WQO 203

OY 280 NADW-GTQ-----WIKDHANAAAAIGKPTILEEFG 308
DB 204 --DWSGTMRSNAASVFASDPDRNTVFSIHMYGVVDATAAEVRDYLNAFVNGNGLPIVVG 261

OY 309 WQ-----TPDRDSVYQTWTQTVRTNCEAGNFWMLAGNVNGQPY-----PN----- 349
DB 262 DOHSDGNPDDEAIIAMAT-AQSLGV-CYLQWS-W--SGNGGGVEYLDWVNGFDNLSLTSGN 316

OY 350 ---YDGFNVVYPSSTATV 364
DB 317 RILYGSNGIAATSRATV 334

Search completed: November 13, 2002, 11:55:03
Job time : 11.6207 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Kun on: November 13, 2002, 10:50:41 : Search time 34.7522 Seconds

(without alignments)
2223.392 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTASGCGFVNLCLPY.....YYPSSTATVLASEALAISTG 375

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREML_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriophage.*
- 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829.5	39.9	666	Q9RIK7	Q9rik7 thermotoga
2	812	39.1	680	Q9RIK9	Q9rik9 thermotoga
3	811	39.0	669	Q9X0V4	Q9x0v4 thermotoga
4	787.5	37.9	694	Q9Z187	Q9z187 bacillus st
5	706.5	34.0	439	Q9P893	Q9p893 agaricus bi
6	706.5	34.0	439	Q9Z401	Q9z401 agaricus bi
7	630	30.3	377	Q00012	Q00012 aspergillus
8	599	28.8	578	Q96V96	Q96v96 orpinomyces
9	596	28.7	437	Q99036	Q99036 trichoderma
10	543.5	26.1	431	Q9FJ23	Q9fj23 arabidopsis
11	537	25.8	408	Q9SG95	Q9sg95 arabidopsis
12	532.5	25.6	416	Q9FT03	Q9ft03 coffea arab
13	530	25.5	395	Q9RVL3	Q9rvl3 lycopersico
14	528	25.4	399	Q93WT4	Q93wt4 lycopersico
15	517.5	24.9	411	Q9FZ29	Q9fz29 arabidopsis
16	516	24.8	414	Q9SG94	Q9sg94 arabidopsis

17	498	24.0	448	10	Q9LZV3	Q9lzv3 arabidopsis
18	495	23.8	431	10	Q9M0H6	Q9m0h6 arabidopsis
19	493	23.7	427	10	Q9FT02	Q9ft02 coffea arab
20	491.5	23.6	439	10	Q8SAY2	Q8say2 oryza sativ
21	483	23.2	445	10	Q94J47	Q94j47 oryza sativ
22	476	22.9	468	10	Q8SAY1	Q8say1 oryza sativ
23	464	22.3	401	10	Q9FU06	Q9fuq6 lycopersico
24	464	22.3	415	10	Q9FZ03	Q9fz03 lycopersico
25	426.5	20.5	369	10	Q48540	Q48540 lycopersico
26	411.5	19.8	403	10	Q9SKU9	Q9sku9 arabidopsis
27	407	19.6	311	10	Q9LW44	Q9lwa4 arabidopsis
28	399.5	19.2	442	16	Q9AA10	Q9aa10 caulobacter
29	255	12.3	218	10	Q93X40	Q93x40 lactuca sat
30	193	9.3	171	10	Q9LKG5	Q9lku5 chlorocella v
31	171.5	8.2	589	16	Q87851	Q87851 streptomyce
32	143	6.9	2817	16	Q97K42	Q97k42 clostridium
33	141.5	6.8	425	2	Q9LAJ3	Q9laj3 clostridium
34	140.5	6.8	649	16	Q8YD80	Q8y80 brucella me
35	140	6.7	367	5	Q8WPJ2	Q8wpj2 mytilus edu
36	139.5	6.7	397	2	Q8RPZ3	Q8rpz3 puenibacilli
37	136.5	6.6	722	16	Q97L56	Q97l56 clostridium
38	135	6.5	660	2	Q9L3J2	Q9l3j2 clostridium
39	134	6.4	482	2	Q9FAC7	Q9fac7 rhodococcus
40	134	6.4	1684	2	Q0J658	Q0j658 unidentified
41	132.5	6.4	727	2	Q9AF65	Q9af65 clavibacter
42	129.5	6.2	741	17	Q87KH5	Q8tkh5 methanobarc
43	128.5	6.2	585	2	Q8VUT3	Q8vut3 pseudomonas
44	127	6.1	425	2	Q93Q07	Q93q07 thermus cal
45	126.5	6.1	745	2	Q9K5C7	Q9k5c7 clavibacter

ALIGNMENTS

RESULT 1

Q9RIK7	ID	Q9RIK7	PRELIMINARY;	PNT;	666 AA.
AC	Q9RIK7				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Beta-mannosidase (Beta-mannanase).				
GN	MANB OR MAN5.				
OS	Thermotoga neapolitana.				
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2337;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=5068;				
RA	Parker K.N., Lam D., Duffaud G., Shead M.A., Mathur E.J., Kelly R.M.;				
RT	"Amino Acid Sequence of beta-mannosidase genes from the				
RT	hyperthermophilic bacteria Thermotoga maritima and Thermotoga				
RT	neapolitana.";				
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DSM 5068;				
RX	MEDLINE=21473892; PubMed=11590605;				
RA	Parker K.N., Chhabra S.R., Lam D., Callen W., Duffaud G.D.,				
RA	Snead M.A., Short J.M., Mathur E.J., Kelly R.M.;				
RT	"Galactomannanases Man2 and Man5 from Thermotoga species: growth				
RT	physiology on galactomannans, gene sequence analysis, and biochemical				
RT	properties of recombinant enzymes.";				
RL	Biotechnol. Bioeng. 75:322-333(2001).				
DR	EMBL; Y17981; CAB56856.1; .				
DR	EMBL; AY033477; AAK53459.1; .				
DR	InterPro: IPR001547; GH5.				
DR	Pfam: PF00150; cellulase_1.				
SQ	SEQUENCE 666 AA; 76695 MW; 7EF986115E4C58CD CRC64;				

Query Match 39.9%; Score 829.5; DB 2; Length 666;
Best Local Similarity 43.5%; Pred. No. 5.5e-50;
Matches 173; Conservative 63; Mismatches 119; Indels 43; Gaps 13;

```

QY      1  APAGFVTTASGCGFVLNCLPYRYCGTNNYYLSTYOSHADVDDVLAKAQMNLNVIRKTHWGFID 60
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     19  ANSDFVKVNGRFLNGEEERFVGSNNYYMHYKSNRMDSVLSESAKAMGVKVLRIWGF-- 76
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY     61  IGSLDGVSPTIDGNKNGCFYQYWDPSGTA---PAYNDGPTGQLQCLDYAJASAAAHGLRVI 117
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     77  --LDGESYCRDN-----TYMHPAGVFGLEPGTNAODGFERLDYTVAKAKELGIKLI 127
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY    118  VVLTDNWKKEFGMDQYDKWYGLBYHNFYDTPRTOQAYKNWNHILLNRVNSITGVTYKND 177
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    128  IVLVNNWDDFCGMNQYVRWFGGIHDDFVENEK1KEEYKVVFLINRVNTYTGVPYREE 187
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY    178  PTIFAWELANPRCVGSGTLPTSGCTCTQATIVANKVDMSAYVKSIDPNHMVSVGDEGF-- 235
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    188  PTIAWELANPRC-----ETDKSGN----TLVEWVEMSAYIKSLDPNHLVAVGDEGFEN 239
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY    236  -YIGSTGCG---WPYNDRPSGDGVNNALLRVKNIDFCTVHLYPNYWGQN-----ADMGTQW 297
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    240  NYEGFRPYGGEAEWAYNGWS-GVMDKRLLE1ETVDFCTFHLVPSHWGVSPENYAOWGAKW 298
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY    288  IKOHIAANAATGKPTILEEFG--WQTP-DRDSVYQYQTWTQVTRTNGEAGNFWMLAGNVNG 344
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    299  IEDHIK1AKEVGPVVLVEYIGPKSPVNRVAIYKILWMDIYVNLGGNGAMPWNMLAGIGEG 358
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY    345  QP-----YPNYGDFNVPYSPSTATVLASE-AIAISTG 375
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    359  WDRDEKGYYPDYDCGRIVNDESEEAKLIREYAKLFSTG 396
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 2
ID      Q9RIK9      PRELIMINARY;      PRT:      680 AA.
AC      Q9RIK9;
DT      01-MAY-2000 (T:EMBLrel. 13, Created)
DT      01-MAY-2000 (T:EMBLrel. 13, Last sequence update)
DT      01-JUN-2002 (T:EMBLrel. 21, Last annotation update)
DE      Beta-mannosidase.
GN      MANB.
OS      Thermotoga maritima.
OC      Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX      NCBI_TaxID:2336;
KW      111
KP      SEQUENCE FROM N.A.
RC      STRAIN=MSB8;
RT      Parker K.N., Lam D., Duffaud G., Sneed M.A., Mathur E.J., Kelly R.M.;
RT      "Amino Acid sequence of beta-mannosidase genes from the
RT      hyperthermophilic bacteria Thermotoga maritima and Thermotoga
RT      neapolitana."
RL      Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR      EMBL: Y17980; CAB56854.1; -
DR      InterPro: IPR004200; Bgal_small_C.
DR      InterPro: IPR001547; GH5.
DK      Pfam: PF02930; Bgal_small_C.1.
DR      Pfam: PF00150; cellulase.1.
SQ      SEQUENCE      680 AA; 77687 MW; 5BA3F6982194D1AA CRC64;

```

Query Match.	39.1%	Score	812:	DB 2:	Length	680;			
Best Local Similarity	42.8%	Pred. No.	9.5e-49;						
Matches	169;	Conservative	61;	Mismatches	121;	Indels	44;	Gaps	13;
Qy	5	FVTASGGQFVLNGLPYRYRGTTNNYYLSQSADVDVLLAKAAQNLSLVIRTWGFDIDCSL	64						
Dbb	34	FKVENCKFALNGKEFRFGISNNYYMIYKSNMGMDSVLESARDMGIKVLRTWGR----	L 88						
Qy	65	DGSVPT IDGNKNNGFYFOYWPSTCAPAYNDG-----PTCQLGLDVIATAAAGHLRUVVL	120						
Dbb	89	DGESYCRDKN-----TYMHPEPCVFVGPEGISNAQSGFERLDVTVAKELGKLVIVL	142						
Qy	121	TNDHKFEFGGDYOYDKWKCLPYHDNFYTDPTQOQAYKWNVNILLNRVNSITGVTVKNDPTI	180						
Dbb	143	VNMDDPFCGMNOYVVRFGTHHDFDYRDEKIKEEYKKYVSFLNVHNATYTCVPYREPTI	202						

Qy	181	FAWELANEPRCVGSGTLPTSGTCTQATIVNVWVQMSAYVKSIDPNHMVSVGDGCF	---	YI	237
Db	203	MAWELANEPRC-----ETDKSGN----	TLVWVKEMSSYIKSLDPNHLVAVDGEFFSNYE	---	254
Qy	238	GSTGSG----	WPYNDPSGDVDNNALLRVKNIIDFGTYHILYPNYNGON----	ADMGCTOWIKD	290
Db	255	GFKPYGGEAWYNGWS-GVDWKKLLSIEVDFGTFHLYPSHWGVSPENYAQNGAKWIED	---	313	
Qy	291	HIANAATIGKPTLEEFQ--WQTP-DRDSVYOTWTQTVRTNGEAGNFMFMLAGNVNGQP-	---	346	
Db	314	HIKIAKEIGKVPWLEEYGIKPSAPNRTATVRLWDLVYDVGGLDGGAMFNLGIGEGSDR	---	373	
Qy	347	-----YPNYDGFNVYPSSTATVLAASE-ALAI	STG	375	
Db	374	DERGYYPDYDGFIRVNDSDSPEAEI	IREYAKLFNTG	408	
RESULT 3					
Q9X0V4					
ID	Q9X0V4	PRELIMINARY;	PRT;	669	AA.
AC	Q9X0V4;				
DT	01-NOV-1999 (TrEMBLrel. 12, Created)				
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)				
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)				
DE	Endo-1,4-beta-mannosidase.				
GN	TM1227.				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae;				
OX	NCBI_TaxID=2336;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSB8 / DSM 3109;				
RX	MEDLINE=99287316; PubMed=10360571;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,				
RA	Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,				
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,				
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;				
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from				
RL	Nature 399:323-329(1999).				
DR	EMBL; AE001779; AAD36302.1; .				
DR	TIGR; TM1227; .				
DR	InterPro; IPR001547; GH_5.				
DR	Pfam; PF00150; cellulase; 1.				
KW	Complete proteome.				
SQ	SEQUENCE 669 AA; 76931 MW; 5659B727305688F5 CRC64;				
	Query Match	39.08;	Score 811;	DB 16;	Length 669;
	Best Local Similarity	42.88;	Pred. No. 1.1e-48;		
	Matches 169;	Conservative 61;	Mismatches 121;	Indels 44;	Gaps
Qy	5	FVTASGQFVLNGLPYRYGGTNNYLSYQSHADVDIVLAKAQAMNLSVIRTWGFI	DIGSL	64	
Db	23	FVKVCKNGKALNGKEFRFGSNYYMYHKSNRMDSVLESARDWGIKVLWGF	----	L	77
Qy	65	DGSVPTIDGNKNGYFYQWDPSTGAPAYNG-----PTGLOGLDYA	TASAAHGLRIVVL	120	
Db	78	DCESYCRDKN-----TYMPEPCVFCVPEGIC	ISNAQSGFERLDY	TVAKAKELG	IKLVVL
Qy	121	TNDMKRFGGMDYDKWGLPYHDHNFYDPTQOAYKNWHLNRYNSITGVY	KKNDPTI	180	
Db	132	VNNWDFDGGMNOYVRFWGGTHDDFYRDEKIKKEYKIVSFLVHNHNTY	YTGVP	REPTI	191
Qy	181	FAWELANEPRCVGSGTLPTSGTCTQATIVNVWVQMSAYVKSIDPNHMVSVGDGCF	---	YI	237
Db	192	MAWELANEPRC-----ETDKSGN----	TLVWVKEMSSYIKSLDPNHLVAVDGEFFSNYE	---	243
Qy	238	GSTGSG----	WPYNDPSGDVDNNALLRVKNIIDFGTYHILYPNYNGON----	ADMGCTOWIKD	290
Db	244	GFKPYGGEAWYNGWS-GVDWKKLLSIEVDFGTFHLYPSHWGVSPENYAQNGAKWIED	---	302	

Query Match 39.0%; Score 811; DB 16; Length 669;
Best Local Similarity 42.8%; Pred. No. 1.1e-48;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

QY 5 FV TASGCGFVLNGLPYRYGGTNNYYLSYQSHADVDJDLAKAQAMNLSVIRTWGFIDIGSL 64
|| || || || || : || || : || || : || || : || || : || || : || || : || || : || ||
Db 23 FYKVENGFALNGKEFREIGSNYYMHYKSNRMIDSVLPSARDMGIKVLRTWGF-----L 77

QY	65	DG	SVPTIDGNKNGFYQYWDPSGAPAYNDG----	PTGLQGLDYATASAAAHGLRIVVL	120		
		II	II	II	II		
DB	78	DG	SEYCRDN-----	TYMHPEGVFGVPEGISNAOSGFERLDY	131		
		II	II	II	II		
QY	121	TND	WKFEFGNDQDKYGLPYHDNFYDPRTOAQYK	WNVHLLNRVNSITGVYTKNDPTI	180		
		II	II	II	II		
DB	132	VNM	DDFGGNGQYVRVFGGTHHDDFYRDEKIK	EYKYVSFLVNHVNTYTGVPREEPTI	191		
		II	II	II	II		
QY	181	FAME	LANEPRCVGSGTLPTSGTCTQATINWVD	OMSAIVKSIDPNHMSVSDGEGF----	237		
		II	II	II	II		
DB	192	MA	ELANEPRC-----	ETDKSGN-----	TLVWYKEMESYIKSLDPNHLVAVGDEG	243	
		II	II	II	II		
QY	238	GST	QGGSG----	WPYNDPSDGDVNNALLRVKNIDFGTY	HLHPYNYWGQN-----	ADMGTOIKD	290
		II	II	II	II	II	II
DB	244	GK	PYCGEAEWAYNGWS-GYDWMKLLIS	ITVDFGTFHLYPSHMGVSPENYA	OWGAKWIED	302	
		II	II	II	II	II	II

13;

5

Mism

svat.

Ma

OS Agaricus bisporus (Common mushroom).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Agaricaceae; Agaricus.
 OX NCBI_TaxID=5341;
 KN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D649; TISSUE=MYCELIUM;
 RX MEDLINE=97177791; PubMed=9025297;
 RA Yague E., Mehak-Zunic M., Morgan L., Wood D.A., Thurston C.F.;
 RT "Expression of CEL2 and CEL4, two proteins from Agaricus bisporus with
 similarity to fungal cellobiohydrolase I and beta-mannanase,
 respectively, is regulated by the carbon source.";
 RT Microbiology 143:239-244(1997).
 RL EMBL: Z50095; CAA90423.1; -.
 DR HSP: P00725; IAZ6.
 DR InterPro: IPR001589; Actbind_actnin.
 DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00734; CBM_1; 1.
 DR ProDom: PD001821; CBD_fungal; 1.
 DR SMART: SM00236; fCHD; 1.
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 DR PROSITE: PS00562; CBD_FUNGAL; 1.
 KW Glycosidase; Hydrolase; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 439 CEL4B MANNANASE.
 SQ SEQUENCE 439 AA: 46940 MW: 7250A2D4A273D522 CRC64;
 Query Match 34.0%; Score 706.5; DB 3; Length 439;
 Best Local Similarity 40.8%; Pred. No. 1.3e-41;
 Matches 154; Conservative 57; Mismatches 131; Indels 35; Gaps 12;
 QY 4 GFVTASGGQFVLNGLPYRGCTNNYY--LSYQSHADVDVLAQAKAMNLSVIRTWGFI 61
 DB 86 GFVKASCTRTLLNGQYTVVGGNSYVWGLTGLSTAMNQAFSDIANAGCTVITWGFNEV 145
 QY 62 GSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRVIVLT 121
 DB 146 TS-----PNGNYIQSW--SGARPTINTGASCLLNFNDVIAAKANGIRLIVALT 192
 QY 122 NDMKEFGMDQYDKWY---GLPYHDNFYTDPRTOQAYKNVNHLLNRVNSITGVTKNDP 178
 DB 193 NNWADYGGHDDVYVQWVGNGOP--HDLFYNTAIDKAFKSYGRAFSR-----YANEP 243
 QY 179 TIFAMELANEPRCVSGTLPTSGCTCTOATIVNWVDQMSAYVKSIDPNHVMVSGDEGFYIG 238
 DB 244 TVNAMELANEPRCKGS--TGTTSGCTCTTTVTNNAKEMSAFIKIDS NHLVAIGDEGFY-N 301
 QY 239 STQGSQWPNYDSDGVDNNALLRVKNIIDFGTHLYPNYWGONAD---WGTQWKIDHANA 295
 DB 302 QPGAPTYPG--SEGVDFAEALAISSVDFAFHSYPEPWQGGADAKWGTQWITDHAASM 360
 QY 296 AATGKPTILEEFGWQTPDRDSVYQVOTVQVTRTNGEAGNFWMLAGNVNGQYPNYDGFNV 355
 DB 361 KRVNKPVILEEFGVTTNQPDY--YAEWFNEIESSGLTGDLIWOAGSHLSTGDTN--DGIYV 418
 QY 356 YPSSSTATVLASALAI 372
 DB 419 YPDGPVYPLVKSHASAM 435
 RESULT 7
 Q00012 PRELIMINARY; PRT: 377 AA.
 AC Q00012;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Mannanase precursor.
 GN MAN1.
 OS Aspergillus aculeatus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5053;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KSM 510;
 RX MEDLINE=95078777; PubMed=7987261;
 RA Christgau S., Kauppinen S., Vind J., Kofoed L.V., Dalboge H.;
 RT "Expression cloning, purification and characterization of a beta-1,4-
 mannanase from Aspergillus aculeatus.";
 RL Biochem. Mol. Biol. Int. 33:917-925(1994).
 DR EMBL: L35487; AAA67426.1; -.
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00150; cellulase; 1.
 KW Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 377 MANNANASE.
 SQ SEQUENCE 377 AA: 41082 MW: 10E6477555BE3CA2 CRC64;
 Query Match 30.3%; Score 630; DB 3; Length 377;
 Best Local Similarity 35.2%; Pred. No. 2.4e-36;
 Matches 132; Conservative 65; Mismatches 142; Indels 36; Gaps 9;
 QY 1 APAGFVTASGGQFVLNGLPYRGCTNNYYLSY-QSHADVDVLAQAKAMNLSVIRTWGFI 59
 DB 28 ATTAFPTSTGLHFTIDGKTGCFAGTNSYWGFLTNDDVDLVMSQLAASDLKILRVWGFN 87
 QY 60 DICS--LDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRV 117
 DB 88 DVNKTPTDGTV-----WYQJHANGTSTJNTGADGLOHLDYVVTSAEYGVKLI 135
 QY 118 VLTNDMKEFGMDQYDKWYGLPYHDNFYTDPRTOQAYKNVNHLLNRVNSITGVTKND 177
 DB 136 INFVNEWTDYGGMOAYVTVAGAAQTDFYNTAIGAAKYNV1KAVVSR-----YSSS 187
 QY 178 PTIFAMELANEPRCVSGTLPTSGCTCTOATIVNWVDQMSAYVKSIDPNHVMVSGDEGFY 237
 DB 188 AATFAMELANEPRCQG-----CDTSVLYNM1SDTSKYIKSLDSKHLVITIGDEGFL 238
 QY 238 GSTQGSQWPNYDSDGVDNNALLRVKNIIDFGTHLYPNYWGONADQWTKIDHIANAA 297
 DB 239 DVSDSGSYPT--VGEGLNFTKNLGISTIDFGTLHLYPDSNGTSDYDNGNGWITAHAAACKA 297
 QY 298 IGKPTILEEFGWQTPDRDSVYQVOTVQVTRTNGEAGNFWMLAGNVNGQYPNYDGFNV 356
 DB 298 VGRKPCLEEYGV-VTSNHCIVESPWQQTAGNATG1SGDLIYQYGTTFPSWGQSPN-DGNTFY 355
 QY 357 YPSSSTATVLASEALA 371
 DB 356 YNTSDFTCCLVTDHVA 370
 RESULT 8
 Q96V96 PRELIMINARY; PRT: 578 AA.
 AC Q96V96;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Mannanase ManA.
 GN MANA.
 OS Orpinomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimasticeae; Orpinomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PC-2;
 RX MEDLINE=21405733; PubMed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL: AF177206; AAL01213.1; -.


```

Db 189 YKDDPTIMAWELNPRCPSPD-----PSGRAIOA-----WITEMAAHVKSLSDRNIIILEAGLE 240
QY 234 GFYIGSTOGSCWPNPDS-DVONNALLRVKNIDFGTYHLYPNYWGONA-----DWCTO 286
Db 241 GFY-GOSSPQSKTLNPPGQGFDTFIANNRIPGIDFVTVHSYPDWEPDSSEQSOMDFLNK 299
QY 287 WIKDHIANAA-AIGKPTILEEF-----GWOTPPDRDSVYOT-----WTOTVTRNGEAGW 333
Db 300 WLDAAHLODAONVLUKHPILIAEPKSKMKKPCYTTPAQRDIVENTVYSKIYGSAAKKGGAAGG 359
QY 334 NFWMLAGNVNGQPNY-DGFNYYPSSTATV 364
Db 360 LFWOLL--VNG--IDNFQDGYGIILSOSSTV 387

RESULT 11
Q9SG95
ID Q9SG95 PRELIMINARY; PRT; 408 AA.
AC Q9SG95;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Putative (1-4)-beta-mannan endohydrolase.
GN T7M13.2.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosidia I; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M.-I., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.V., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T7M13 genomic sequence.";
DR EMBL; AC011708; AAF19559.1; -.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW hydrolase.
SQ SEQUENCE 408 AA; 45376 MW; 86C48FA7D311BA15 CRC64;

Query Match 25.8%; Score 537; DB 10; Length 408;
Best Local Similarity 33.3%; Pred. No. 8.7e-30;
Matches 135; Conservative 74; Mismatches 128; Indels 68; Gaps 16;

QY 1 APA-GFVTASGGQFVLNGLPYRYGGTNNYLSYQ-----SHADVDVLAQAQAMNLSVIR 54
Db 24 APSDGFVSRNGVQF'LNGKPFYANGFNAYWLAYEATDPTTRFKITNVFQNAISLGLTIAR 83
QY 55 TWGCFIDIGSDGVSPTIDGNKNGFYFOYWDPTGAPAYNDGPTGLOGLDYAIASAAHGL 114
Db 84 TWGFRD-GAIYRALQAPGS-----YDEQT-----FQGLDFVIAEAKRIGI 123
QY 115 RVIVLTLNDWKEFGMDQYDKWY-----LPYHDNFYTDPTQOAYKNWVHLLNRVNSI 169
Db 124 KLILLVNNWDDYGGKKQYVDWARSKEVSVSSNDDEYRNPVIFKDFYKNHVKTVLNRVNTF 193
QY 170 TGVTYKNDPTIFAWELANEPRCVSGTSLPTSGTCTQATIVNVVDQMSAYVKSIDPNHVS 229
Db 184 TKVAYKDEPAINAWQLMNEPRC-----GVDKSG-----KTLMDMINEMAPFVKSVDPNHLS 235
QY 230 VGDEGYIGST---QSGGWPYNDPSDGVNDNALLRVKNIDFGTYHLYPNYWGONADWGT- 285
Db 236 TGHEGYGSSSPERKNSLPVSNANTVGAQDFIAHNHIDAIDFASMHCGSGLWFLRQNSR 295
QY 286 -----QWIKDHIANAAI-GKPTILEFGW--QTP-----DRDSVYOT-----WTOTVRT 327
Db 296 LAFIKWLEGHIEDAQNILKKPVILAEFLGSDTPRYTILANRDCGVTTTYYIIYASQKG 355

```

```

QY 328 NGEAGNFWMLAGNVNGQPNYVDG-----FNYYPSSTATVLASEA 369
Db 356 GSAAGALFW-----EVISEGMSNFAGPSSIIILSDKSTVNTIIEHA 396

RESULT 12
Q9FT03
ID Q9FT03 PRELIMINARY; PRT; 416 AA.
AC Q9FT03;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE (1-4)-beta-mannan endohydrolase precursor (EC 3.2.1.78).
GN MANB.
OS Coffea arabica (Coffee).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Gentianales; Rubiaceae; Ixoroideae; Coffeae;
OC Coffea.
OX NCBI_TaxID=13443;
RN [1]
RC TISSUE=GRAIN;
RA Marraccini P., Rogers W.J., Allard C., Andre M.L., Cailliet V.,
RA Lausanne F., Michaux S.;
RT "Molecular and biochemical characterization of endo-b-mannanase from
RT germinating coffee (Coffea arabica) grains.";
RL Planta 213:296-308(2001).
DR EMBL; AJ278996; CAC08442.1; -.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Glycosidase: Hydrolase; Signal.
FT SIGNAL 1 40
FT CHAIN 41 416 ENDO BETA MANNANASE.
SQ SEQUENCE 416 AA; 46824 MW; E484CCCAFC674D66 CRC64;

Query Match 25.6%; Score 532.5; DB 10; Length 416;
Best Local Similarity 32.3%; Pred. No. 1.8e-29;
Matches 134; Conservative 65; Mismatches 127; Indels 89; Gaps 19;

QY 5 FVTASGGQFVLNGLPYRYGGTNNYLSY-----QSHADVDVLAQAQAMNLSVIRTWGFI 59
Db 44 FVKTRGTEFVNGRPLVNGFNAYWLAYMASDPSTRTKVSTTFOASKYGNNAARTWAFS 103
QY 60 DIGSLDGSVPTIDGNKNGFYFOYWDPTGAPAYNDGPTGLOGLDYAIASAAHGLRIVV 119
Db 104 DGG-----YRALQSPG--SYNEDM--FKGLDFVYSEAKKYGIHLILT 142
QY 120 LTNDWKEFGGMDQYDKW-----YCLPYHDNFYTDPTQOAYKNWVHLLNRVNSITGVY 174
Db 143 LVNWEYGSGKGKQVQWARDQGYLNNDDDFDTDPVIRGYFKNHKIVTLRINSITGLAY 202
QY 175 KNDPTIFAWELANEPRCVSGTSLPTSGTCTQATIVNVVDQMSAYVKSIDPNHIVSVGDEG 234
Db 203 KDPPTIFAWELMNEPRCSD-----LSCKAIQ-----DWISEMATHVKSIDSDDLIDGLEG 254
QY 235 FYTGSTQGSQWP-----YNDPSDGVNDNALLRVKNIDFGTYHLYPNYWGONADWGT 286
Db 255 FY-----GESVPQKEYNPGYQVGTDFISNN---RIVQVDFATIHLYPDQWVPNSNDETQ 306
QY 287 -----WIKDHIANAA-AIGKPTILEEF-----GWOTPPDRDS-----VYOT--WTOTVR 326
Db 307 AQFVDRWIKHIEDQSKYLLLEKPLLLTEFGKSSRSPGVQAKRDAYLSHIVDTIYACAA 366
QY 327 TNGEAGN-FWMLAGNVNGQPNY-----DCFNVYPS--STATVLASEAIAIST 374
Db 367 GGCVCGNLPMQWA-----PGMESWGQGYEIVLEENPSTVGVIAQSNRLSS 414

RESULT 13
QBRVL3

```

ID Q8RVL3 PRELIMINARY: PRT: 395 AA.
AC Q8RVL3;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE (inactive endo-beta-mannanase)
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WALTER. TISSUE=RIPENING FRUIT PERICARP;
RA Bourgault R., Bewley J.D.;
RT "A cDNA encoding an inactive endo-beta-mannanase expressed in ripening
RT tomato fruit of the cultivar Walter."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046589; AAK97759.1; -
SQ SEQUENCE 395 AA: 44837 MW: 90A0DCA6ACD45F9B CRC64;
Query Match 25.5% Score 530; DB 10; Length 395;
Best Local Similarity 34.6%; Pred. No. 2.6e-29;
Matches 140; Conservative 55; Mismatches 132; Indels 78; Gaps 18;
QY 5 FVTASGCGQFVNLGPRYCGTNNYLSY-----QSHADVDVLAQAAMNLSVIRTWGFI 59
DB 32 FVYTNCTHIFALNGKSLYINGFNAYMLYIAYDPSTRIKVTNTFQQAASKYKMNVAWTF 91
QY 60 DIGSLDGSVPTIDGKNGFYFOWDSTGAP-AYNDGPTGLOGLDYALASAAHGLRVIV 118
DB 92 HGS-----RPLQAPGVNE--QMFQGLDPVISEAKKYGIHLIM 129
QY 119 VLTNDMKEFGMDQYDKW-----YGLPYHDNFYDTPRTOQAYKNVHLLNRVNSITGVT 173
DB 130 SLVNNDAFCGKKQYVEMAVQRCGLTSDDDFTNPMVKGYKNNVKKVLTTRVNTITKVA 189
QY 174 YKNDPTIFAWELANPRCVSGTLPTSGTCTQATIVNWDQMSAYVKSIDPNHMVSGDE 233
DB 190 YKDDPTILSWELINEPRCPSPD-----LSG-----KTFQNWVLEMAGYLKSIDSNHLLEIGLE 241
QY 234 GFVYIGTSGGWPNDPS-----DGVNNALLRVKNIIDFCTHYLYPNW-----GONAD-- 282
DB 242 GFYGNDR-----QYNPSNYIFGTNFIISNN--QVQIDFTTHMYPNQWLPGLTQEAQDK 294
QY 283 WGTQWTKHIAANAAATGKPTILEEFGWOT-----PDRDS-----VYQWTQTVRTNGEA 331
DB 295 WASQWIOVHIDSKMLKPLLIAEFCKSTKTCYTVAKRDNYFEKIYGTIFNCAKSGGPC 354
QY 332 GWN-FWMLAGNVGQPPNY-DGFNVYPS--STATVLAASEALAI 372
DB 355 GUGLFW-----QVLQGMSSFDGQYVVLQESPSRVSILLQSLAL 395
RESULT 14
Q93WT4
ID Q93WT4 PRELIMINARY: PRT: 399 AA.
AC Q93WT4;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Endo-beta-mannanase (EC 3.2.1.78).
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. TRUST. TISSUE=RIPENING FRUIT PERICARP;
RA Bourgault R., Bewley J.D.;
RT "A cDNA encoding an endo-beta-mannanase expressed in ripening tomato
RT fruit of the cultivar Trust;"

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046588; AAK97760.1; -
DR InterPro: IPR001547; GH.5.
DR Pfam: PF00150; cellulase.1.
KW Glycosidase; hydrolase.
SQ SEQUENCE 399 AA: 45338 MW: 805A7D7866E5FF6B CRC64;
Query Match 25.4% Score 528; DB 10; Length 399;
Best Local Similarity 34.5%; Pred. No. 3.6e-29;
Matches 140; Conservative 54; Mismatches 134; Indels 78; Gaps 18;
QY 5 FVTASGCGQFVNLGPRYCGTNNYLSY-----QSHADVDVLAQAAMNLSVIRTWGFI 59
DB 32 FVYTDGTHFALNGKSLYINGFNAYMLYIAYDPSTRIKVTNTFQQAASKYKMNVAWTF 91
QY 60 DIGSLDGSVPTIDGKNGFYFOWDSTGAP-AYNDGPTGLOGLDYALASAAHGLRVIV 118
DB 92 HGS-----RPLQAPGVNE--QMFQGLDPVISEAKKYGIHLIM 129
QY 119 VLTNDMKEFGMDQYDKW-----YGLPYHDNFYDTPRTOQAYKNVHLLNRVNSITGVT 173
DB 130 SLVNNDAFCGKKQYVEMAVQRCGLTSDDDFTNPMVKGYKNNVKKVLTTRVNTITKVA 189
QY 174 YKNDPTIFAWELANPRCVSGTLPTSGTCTQATIVNWDQMSAYVKSIDPNHMVSGDE 233
DB 190 YKDDPTILSWELINEPRCPSPD-----LSG-----KTFQNWVLEMAGYLKSIDSNHLLEIGLE 241
QY 234 GFVYIGTSGGWPNDPS-----DGVNNALLRVKNIIDFCTHYLYPNW-----GONAD-- 282
DB 242 GFYGNDR-----QYNPSNYIFGTNFIISNN--QVQIDFTTHMYPNQWLPGLTQEAQDK 294
QY 283 WGTQWTKHIAANAAATGKPTILEEFGWOT-----PDRDS-----VYQWTQTVRTNGEA 331
DB 295 WASQWIOVHIDSKMLKPLLIAEFCKSTKTCYTVAKRDNYFEKIYGTIFNCAKSGGPC 354
QY 332 GWN-FWMLAGNVGQPPNY-DGFNVYPS--STATVLAASEALAI 373
DB 355 GUGLFW-----QVLQGMSSFDGQYVVLQESPSRVSILLQSLRLS 396
RESULT 15
Q9FZ29
ID Q9FZ29 PRELIMINARY: PRT: 411 AA.
AC Q9FZ29;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE T6A9.1 protein (1-4).
GN T6A9.1 OR ATIG02310.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chlou J., Choi E., Gonzalez A.,
RA Howng B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharsy N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC064879; AAG00883.1; -

```
DR EMBL: AY081352; AAL91241.1; -.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
KW Hydrolase.
SQ SEQUENCE 411 AA: 46290 MW: 872ECB54C4D7F218 CRC64;

Query Match      24.9%; Score 517.5; DB 10; Length 411;
Best Local Similarity 32.3%; Pred. No. 2e-28;
Matches 130; Conservative 69; Mismatches 132; Indels 71; Gaps 14;

Qy 4 GFVTASGQFVLNGLPYRYGGTNNYYL-----SYQSHADVDDVLAKAAQAMNLSVIRTW 56
Db 28 GFVGHNGTQFVLNGEQVYLNGFNAYMMTTAAADTASKGRATVTTALRQASAYGMNVARIW 87
Qy 57 GFIDIGSLDGSVPTIDGNKNGFYFOYWPDPSTGAP--AYNDGPTGLOGLDYAIASAAAHLR 115
Db 88 GF-----NEGDYI-----PLQISPGSYSEDV--FKGLDPVYVEAGRNIK 125
Qy 116 VIVVLTDNWKFGMDQYDKWYGLPYHDNFTDPRTOQAYKNVNHLLNRVNSITGVYTK 175
Db 126 LIISLVNPFEDYGGKKYVENAGLDEPDEFYNSAVKQFYKNHVKTVLTKNTITGRMYK 185
Qy 176 NDPTIFAWELANEPKCVSGSLPTSGCTOATIV--NWVDQMSAYVKSIDPNHMSVGDGEG 234
Db 186 DDPTIFSWELINEPRCND-----TASNILODWKEMASYKSDSNHLLLEIGLEG 236
Qy 235 FYIGSTQSGWPYN-----DPSGVDNALLRVKNIDFGTYHLYPNYW-----GON 280
Db 237 FY-GESIPERTVYNGGRVLTGDTITNN---QIPDIDFATIHYPDSWLPQSSRTGEQ 292
Qy 281 ADMGTQWIKDHIANA-AAIGKPTILEEF-----GWOTPDSDSVYQTWTQTVRTNGEAG 332
Db 293 DTFVDKRWICAHIEDCDNITKKPLLITEFGKSKKYPGFSLEKRNKEFKQYVDVIYDSARAG 352
Qy 333 WN-----FWMLACNVNGOPYPNYDGFNVYPPSSSTATVLASEA 369
Db 353 GSCTGGVFWQLTTRTGLLGDCGTEVFMQAGPNTTAQLIADQS 394
```

Search completed: November 13, 2002, 11:54:09
Job time : 36.7522 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:49:46 ; Search time 11.5841 Seconds
(without alignments)
952.482 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFVTASGQFVLNGLPY.....YPPSSTATVLAASEALAISTG 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents .AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	812	39.1	663	4	US-09-134-078-61 Sequence 61, Appl
2	812	39.1	680	4	US-09-134-078-25 Sequence 25, Appl
3	630	30.3	377	1	US-08-525-697-2 Sequence 2, Appl
4	179.5	8.6	358	1	US-08-604-913B-11 Sequence 11, Appl
5	179.5	8.6	521	1	US-08-276-213-3 Sequence 3, Appl
6	124.5	6.0	711	4	US-08-961-083-82 Sequence 82, Appl
7	123.5	5.9	430	2	US-08-924-440-2 Sequence 2, Appl
8	121.5	5.8	429	1	US-08-745-977-4 Sequence 4, Appl
9	121.5	5.8	429	3	US-09-040-699A-4 Sequence 4, Appl
10	118	5.7	551	2	US-09-033-537A-1 Sequence 1, Appl
11	117.5	5.7	956	4	US-09-134-078-63 Sequence 63, Appl
12	110.5	5.3	501	4	US-09-465-519-4 Sequence 4, Appl
13	109	5.2	296	1	US-08-507-431-6 Sequence 6, Appl
14	109	5.2	296	3	US-09-116-622-6 Sequence 6, Appl
15	109	5.2	296	4	US-09-219-277-6 Sequence 6, Appl
16	109	5.2	296	4	US-09-599-661-6 Sequence 6, Appl
17	109	5.2	467	2	US-08-727-548-2 Sequence 2, Appl
18	109	5.2	467	4	US-08-945-574-1 Sequence 1, Appl
19	108.5	5.2	634	4	US-09-295-744A-2 Sequence 2, Appl
20	108.5	5.2	10182	4	US-09-134-001C-3159 Sequence 3159, Ap
21	106.5	5.1	616	4	US-09-136-574A-47 Sequence 47, Appl
22	106.5	5.1	879	1	US-08-306-546C-2 Sequence 2, Appl
23	106.5	5.1	879	2	US-08-530-524A-2 Sequence 2, Appl
24	106.5	5.1	1426	4	US-09-136-574A-43 Sequence 43, Appl
25	106	5.1	400	2	US-08-713-298B-2 Sequence 2, Appl
26	106	5.1	400	2	US-08-870-180B-2 Sequence 2, Appl
27	106	5.1	400	3	US-08-814-052-4 Sequence 4, Appl

28	106	5.1	400	3	US-08-812-829-4	Sequence 4, Appl
29	106	5.1	400	4	US-09-226-529-2	Sequence 2, Appl
30	105	5.1	462	2	US-08-870-180B-13	Sequence 13, Appl
31	105	5.1	462	4	US-09-226-529-13	Sequence 13, Appl
32	104.5	5.0	501	4	US-09-465-519-2	Sequence 2, Appl
33	104	5.0	485	4	US-09-291-023A-15	Sequence 15, Appl
34	104	5.0	485	4	US-09-291-023A-2	Sequence 2, Appl
35	104	5.0	603	4	US-09-149-727-6	Sequence 6, Appl
36	104	5.0	832	3	US-08-630-820-7	Sequence 7, Appl
37	103.5	5.0	600	6	5268463-2	Patent No. 5268463
38	103.5	5.0	602	2	US-08-882-704A-5	Sequence 5, Appl
39	103.5	5.0	602	4	US-09-151-957-5	Sequence 5, Appl
40	103.5	5.0	602	6	5432081-2	Patent No. 5432081
41	103.5	5.0	1242	4	US-09-488-270A-2	Sequence 2, Appl
42	102.5	4.9	461	4	US-09-134-001C-5311	Sequence 5311, Ap
43	102.5	4.9	485	4	US-09-291-023A-4	Sequence 4, Appl
44	102.5	4.9	485	4	US-09-291-023A-14	Sequence 14, Appl
45	102.5	4.9	613	4	US-09-149-727-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-134-078-61
; Sequence 61, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
US-09-134-078-61

Query Match 39.1% Score 812; DB 4; Length 663;
Best Local Similarity 42.8%; Pred.No. 4.8e-67;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;
Qy 5 FVTASGGQFVLNGLPYRGYGGTNNYLSYQSHADVDVLAKAQMNLVIRTWGTIDIGSL 64

```

Db 17 FVKVNGKFKALNGKEFRFGSNYYMHYKSNMGIDSVLESARDMGKVLRIWGF-----L 71
Qy 65 DGSVPTIDGNKNGFYFOYWDPSGTAPAYNDG----PTGLOGLDYAIASAAHGLRIVL 120
Db 72 DGSVCRDN-----TYMHPEPGVFGVPEGISNAQSGFERLDYTVAKAKEIGIKLVIVL 125
Qy 121 TNDWKEFGGMDQYDKWYGLPYHDNEYTDPRTOQAYKNWYNHLLNRVNSITGVYTKNDPTI 180
Db 126 VNNWDDFGGMMQYVRWFGCTHDDFYRDEKIKEEYKYYVFLVNVHTYTVGPYREPTI 185
Qy 181 FAWELANPRCVCSTLPTSGTCTOATIVNVWDOMSAYVKSIDPNHMSVGDGEGF---YI 237
Db 186 MAWELANPRC-----ETDKSGN---TLVEWYKEMSSYIKSLDPNHLVAVGDEGFFSNYE 237
Qy 238 GSTQSG---WPYNDPSCVDNNALLRVKNIDFGTYHLYPNYWGON---ADMGTQWIKD 290
Db 238 GFKPYGGEAEWAYNGWS-CVDWKKLLSIETVDFGTHLYPSHMGVSPENYAQWCAKWI 296
Qy 291 HIANAAAIKGPITILEEFG--WQTP-DRDSVYQWTQTVRTNGEAGNFWMLAGNYNGQP- 346
Db 297 HIKIAKEIGKPVVLEEGYGIPIKSPAPVNRNTAIYRLWDLVYDLGCGDGMFWMLAGIGESDR 356
Qy 347 -----YPNYDGNVYPPSSTATVLAASE-ALAI 375
Db 357 DERGYYPDYGFRIVNDDSPAEALIREYAKLFNTG 391

```

RESULT 2

```

US-09-134-078-25
: Sequence 25, Application US/09134078
: Patent No. 6368844
: GENERAL INFORMATION:
: APPLICANT: Bylina, Edward J.
: TITLE OF INVENTION: GLYCOSIDASE ENZYMES
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gray Cary Ware & Freidenrich LLP
: STREET: 4365 Executive Drive, Suite 1600
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09134.078
: FILING DATE: 13-AUG-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/949,026
: FILING DATE: 10-OCT-1997
: APPLICATION NUMBER: 60/056,916
: FILING DATE: 06-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 858/677-1456
: TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 680 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
US-09-134-078-25

```

```

Query Match 39.1%; Score 812; DB 4; Length 680;
Best Local Similarity 42.8%; Pred. No. 5e-67;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;
Qy 5 FYTASGGQVPLNGLPYRYGCTNNYYLSYOSHADVDDVLAKAQAAMNLSVIRTWGTFIDIGSL 64
Db 34 FYKVENGFALNGKEFRFGSNYYMHYKSNMGIDSVLESARDMGKVLRIWGF-----L 88
Qy 65 DGSVPTIDGNKNGFYFOYWDPSGTAPAYNDG----PTGLOGLDYAIASAAHGLRIVL 120
Db 89 DGSVCRDN-----TYMHPEPGVFGVPEGISNAQSGFERLDYTVAKAKEIGIKLVIVL 142
Qy 121 TNDWKEFGGMDQYDKWYGLPYHDNEYTDPRTOQAYKNWYNHLLNRVNSITGVYTKNDPTI 180
Db 143 VNNWDDFGGMMQYVRWFGCTHDDFYRDEKIKEEYKYYVFLVNVHTYTVGPYREPTI 202
Qy 181 FAWELANPRCVCSTLPTSGTCTOATIVNVWDOMSAYVKSIDPNHMSVGDGEGF---YI 237
Db 203 MAWELANPRC-----ETDKSGN---TLVEWYKEMSSYIKSLDPNHLVAVGDEGFFSNYE 254
Qy 238 GSTQSG---WPYNDPSCVDNNALLRVKNIDFGTYHLYPNYWGON---ADMGTQWIKD 290
Db 255 GFKPYGGEAEWAYNGWS-CVDWKKLLSIETVDFGTHLYPSHMGVSPENYAQWCAKWI 313
Qy 291 HIANAAAIKGPITILEEFG--WQTP-DRDSVYQWTQTVRTNGEAGNFWMLAGNYNGQP- 346
Db 314 HIKIAKEIGKPVVLEEGYGIPIKSPAPVNRNTAIYRLWDLVYDLGCGDGMFWMLAGIGESDR 373
Qy 347 -----YPNYDGNVYPPSSTATVLAASE-ALAI 375
Db 374 DERGYYPDYGFRIVNDDSPAEALIREYAKLFNTG 408

```

RESULT 3

```

US-08-525-697-2
: Sequence 2, Application US/08525697
: Patent No. 5795764
: GENERAL INFORMATION:
: APPLICANT: Christgau, Stephan
: APPLICANT: Andersen, Lene N
: APPLICANT: Kauppinen, Sakari
: APPLICANT: Heldt-Hansen, Hans P
: APPLICANT: Dalboege, Henrik
: TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5795764o No. 5795764disk of No. 5795764th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/525.697
: FILING DATE: 21-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Harrington, James J.
: REGISTRATION NUMBER: 38,711
: REFERENCE/DOCKET NUMBER: 4004.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK 0486/93
: FILING DATE: 30-APR-1993
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 2:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-525-697-2

Query Match 30.3%; Score 630; DB 1; Length 377;
Best Local Similarity 35.2%; Pred. No. 1.7e-50;
Matches 132; Conservative 65; Mismatches 142; Indels 36; Gaps 9;

QY 1 APAGFVTSAGGQFV-LNGLPYRYGGTN-QSHADVDDVLAKAQAMNLSVIRTWGFI 59
DB 28 ATAPFSTSLUFTIDGKTCYFAGTNSYWGIFLTNDDVDLVMSQLAASDLKILRVWGFN 87
QY 60 DIGS--LDGSVPTIDGNKNGFYFQYWDPSGTGAPAYNDGPTGLQGLDYAIASAAHGLRVI 117
DB 88 DVNTRKPTDCTV-----WYQLHANGTSTINTGADGLQRLDYVVTSAEKYCVKLI 135
QY 118 VVLTNDWKKEFGMDQYKKYGLPYHDNFYDPTQOAYKNWVHLLNKNVNSITGVYTKND 177
DB 136 INFVNEWTGYGMOAYVTAAGAAQDFYNTAIAQAAKYNIKAVYSR-----YSSS 187
QY 178 PTIFAWELANPRCVSGTLPSTSGTCTOATIVNWVDMSAYVKSIDPNHMSVGDGEFYI 237
DB 188 AAIFAWELANPRCG-----CDTSVLYNWIISDTSKYIKSLDSKHLVITIGDEGFG 238
QY 238 GSTQSGWPNPDSQVDNALLRVKNIDFGTYHLIPYNTWGNADWGTQWIKDHIANAAA 297
DB 239 DVDSGCSYPT-YGEGLNFTKNLGISTIDFGTLHLYPDSWGTSDWNGWITAHAAACKA 297
QY 298 ICKPTILEEFGHQTDRDSVYQWTQTV-RTNCEAGWFWMLAGNVNGQPYPNYDGFNY 356
DB 298 VKRPLCLEBYG-VTSHNICAVESPMOQTAGNATGISDGLYMQYGTTFESWGQSPN-DGNTFY 355
QY 357 YPSSTATVLASEALA 371
DB 356 YNTSDPTCLVTDHVA 370

RESULT 4
US-08-604-913B-11
Sequence 11, Application US/08604913B
Patent No. 5712142

GENERAL INFORMATION:
APPLICANT: Adney, William S.
APPLICANT: Thomas, Steven R.
APPLICANT: Himmel, Michael E.
APPLICANT: Baker, John O.
APPLICANT: Chou, Yat-Chen
TITLE OF INVENTION: METHOD FOR INCREASING
THERMOSTABILITY IN CELLULASE ENZYMES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: U.S.A.
ZIP: 80401-3393

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBC PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II (DOS) text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,913B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,213
FILING DATE: 15-07-1994
ATTORNEY/AGENT INFORMATION:
NAME: Edna M. O'Connor

REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: 95-56
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/384-7573
TELEFAX: 303/384-7499

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
FEATURE:
NAME/KEY: EI-CAT
US-08-604-913B-11

Query Match 8.6%; Score 179.5; DB 1; Length 358;
Best Local Similarity 24.0%; Pred. No. 1.2e-08;
Matches 92; Conservative 50; Mismatches 129; Indels 113; Gaps 24;

QY 1 APAGFVTSAGGQFV-LNGLPYRYGGTN-----NYVLSYQSHADVDDVLAKAQAMNLSV 52
DB 1 AGGYWHITSGREILDANNVVPRIAGINWFGFETCNVYVHGLMSRDYRSMLDQIKSLGYNT 60
QY 53 IRTWGFIDIGSLDGSVPTIDGNKNGFYFQYWDPSGTGAPAYNDGPTGLQGLDYAIASAAH 112
DB 61 IRL-PYSDDDILKPTMP-----NSINFQMNQD-----LOGLTSLQVMDKIVAYAGO 107
QY 113 GLRIVVLTNDWKKEFGMDQYKKYGLPYHDNFYDPTQOAYKNWVHLLNKNVNSITGV 172
DB 108 GLRII--LORHRPDCSG--QSALWY-----TSSVSEATWISDL-----QALAQ 146
QY 173 TYKNDPTIFAWELANPR---CVSGTLPSTSGTCTOATIVNW---VDMASAYVKSIDPNH 226
DB 147 RYKGNPTVVGFDLHNEPHDPACHGCCD-PS-----IDWRLAERAGNAVLSVNP 196
QY 227 MYSV-----GDEGFIYSTGCSG-WPYNDPSQVDNALLRVKN-----IDFGTYHL 272
DB 197 LIFVEGVQSYNGDSYWMGNGNLGAGQVP-----VVLNVPNRLVYSAHDYAT-SV 244
QY 273 YPNWYQONAD-----WCTQW---IKDHIAANAAAIKPTILEEFGHQTDRDSVYQWTW 321
DB 245 YPOTWFSDDPTFPNMPGIMKNKNGYLFNQNIA-----PVMJGEPG--TTLOSTTDOTW 295
QY 322 TOT-----VRTNGEAG-----WNFW 336
DB 296 LKTLVOYLRLPTAGYGADSFQWTFW 319

RESULT 5
US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Himmel, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE EI ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213

```

; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)231-1000
; TELEFAX: (303)231-1098
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-276-213-3

Query Match      8.0%  Score 179.5; DB 1; Length 521;
Best Local Similarity 24.0%  Pred. No. 2e-08;
Matches 92; Conservative 50; Mismatches 129; Indels 113; Gaps 24;

QY 1 APAGFVTASGGQFV-LNGLPYRGCTN-----NYLSYQSHADVDVLAKAQMNLVS 52
Db 1 AGGGYHTSREILDANNVPRIAGINMFGFETCNVYVHGLWSRDYRSLDLOIKSLGYNT 60
QY 53 IRTWGETIDSLDGSVPTIDGNKGFYFOYWDPTGAPAYNDGPTGLOGLDYATASAAH 112
Db 61 IRL-PYSDDLKPGTMP-----NSINFQMNQD-----LQGLTSLOVMDKIVAYAGQI 107
QY 113 GLRVIVVLTNDWKEFGGMDQYDKWGLPYLHDNFYTPDRTQOAYKNVNHLLNRVNSITGV 172
Db 108 GLRLI--LDRHRPDCS--QSALWY-----TSSVSEATWISDL-----QALAQ 146
QY 173 TYKNDPITFAWELANER---CYSGSLPTSGTCTQATVNW---VDQMSAYKSIDPNH 226
Db 147 KYKGNPTVVGFDLHNEPHDPACWCGGD-PS-----IDWRLAARAGNAVLVSVPNL 196
QY 227 MVSV-----GDEGFYIGSTQSG-WPYNDPDSQVDNNALLRVKN-----IDFGYHL 272
Db 197 LIFVEGVQSYNGDSYWWGNLOGAGQY-----VVLNVPRLVYSAHDYAT-SV 244
QY 273 YPNWQONAD-----WGTQM---TKDHIANAAAIKGPITILEEFGWQTPDRDSVYQTV 321
Db 245 YPQTFSDPTFPNMMPGIWNKNWGLFNLQIA-----PWLCEFG--TTLQSTTDQTV 295
QY 322 TOT-----VRTNGEAG-----WNFW 336
Db 296 LKTLVQLRPTAQYGADSFQWTFW 319

RESULT 6
US-08-961-083-82
; Sequence 82, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2

```

```

; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-82

Query Match      6.0%  Score 124.5; DB 4; Length 711;
Best Local Similarity 19.7%  Pred. No. 0.0041;
Matches 72; Conservative 48; Mismatches 126; Indels 119; Gaps 17;

QY 65 DGSVPTIDGNKN--GFYFOYWDPTG-----APAYNDGPTGLO----- 100
Db 371 DGQL--VDAKKDLFGYRYHWTNPFNEGFSNLGERIKFHGCVSLHHDHGALGAENEYKAEYRK 428
QY 101 -----GLDYATASAAHGLRVIVLTNDWKEFGGMDQYDKWGLPYH 142
Db 429 LKQMKEMCVNSIRTHNPASEQTLOIAAELCLLVQEEAFDTW--YGGKKPYD--YG-RFF 483
QY 143 DNFTDPTQOAYKNVNHLLNRVNSITGVTYKNDPTIFAWELANEPKRCVSGTLPTSGT 202
Db 484 EKDATHEARKGER-WSDFDLRTWVE-----RGKNPAIFMMSIGNE---ICEANGDAHSL 535
QY 203 CTQATIVNWDOMSAYVKSIDPNHMSVGDECFYIGSTQSGWPNYNDPSDGV-----DN 256
Db 536 AT-----VKRLVKVTKDVDTRYTMGADKFRFGNGSGGHEKIADELDAVGFNYSEDN 588
QY 257 NALLRVKNIDF-----GTVHL-----YPNY-----WQON 280
Db 589 YKALRAKHPKWLIVYGETSSATRTGSSYRPERELKHNSGPERNYEQSDYCNDRVGCWKT 648
QY 281 ADWGTQWIKDHIAAAA-----ICKPTILEEFGWQTPDRDSVYQTVQTVRTNGRAG 332
Db 649 ATASWTFDRONAGYAGOFIWTGTDYICEPT-----PWHNQNQTPVKSSYFGIVDTAGIPK 703
QY 333 WNFWM 337
Db 704 HDFYL 708

RESULT 7
US-08-924-440-2
; Sequence 2, Application US/08924440
; Patent No. 5871550
; GENERAL INFORMATION:
; APPLICANT: Frits et al.
; TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULOSE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

```

[illegible]

```

RESULT 9
US-09-040-699A-4
: Sequence 4, Application US/09040699A
: Patent No. 6022723
: GENERAL INFORMATION:
: APPLICANT: Kofod, Lene V.
: APPLICANT: Andersen, Lene N.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Christgau, Stephan
: APPLICANT: Dalboe, Henrik

```

APPLICANT: Olsen, Hans S.
TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglucanase
TITLE OF INVENTION: Activity
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: No. 60227230 No. 6022723disk of No. 6022723th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-0401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,699A
FILING DATE: 18-March-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 4175.214-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-040-699A-4

Query Match 5.8%; Score 121.5; DB 3; Length 429;
Best Local Similarity 22.6%; Pred. No. 0.0038;
Matches 85; Conservative 46; Mismatches 142; Indels 103; Gaps 18;
QY 27 NYLVSQSHADVDVLAQAKAMLSVIRTWGFI DGLSGDSVPTIDGKNKNGFYQYWD-- 84
Db 106 NHRDWINPATVQSV-----HDVGLNTR-----IPIG-----YMSYN 138
QY 85 --PSTGAPAYNDGPTGLQGLDYAIAASAAHGLRVIVVLTNDWKEFGMDQYDKWYGL-PY 141
Db 139 AIVDTASEPADGNLQPLDVAVQKAADLGIVYII-----DLHGAPCGQQQDAFTGQNP 194
QY 142 HDNFYDPTQQAYKKNVHLLNRVNSITGVYTKNDPTIFAWELANEP--RCVGSGLPT 199
Db 195 PAGFYNTDYGRAEK-WLSWMTNRH--TNPAYS---TVGMIEVLNEPVRHDDGGGRYP 248
QY 200 SGTCTQATIVNWDQMSAYVKSIDPNHMSVGDGEF-----YIGSTQSGWP 246
Db 249 PG-----QDPSPVQTYTPGALKAVRDAEALNVPKLLHVQFMSKNWSDGP 296
QY 247 -----YNDPSDGVNALLRVKNIDEGTYHLYPNYQGNADWGTQWIKDHIA----- 293
Db 297 RSNAVKNDPMVGFDD-----HNYIGFALQNT--GDOYSLMHSACTDSRVV 340
QY 294 ---NAAAIKPTILEERGMOTPDRDSVYQWTQTV-RTNCEAGNFWMLAGNVGPYP- 348
Db 341 SGQDFALTGWSMTSGADWHDGNFFTKFTTAQOOLYESPGMDGMYIWTWKTENDPRWTY 400
QY 349 NYDGFNVPYPSSTATV 364
Db 401 SYATYLNLYIPTNAAL 416

RESULT 10
US-09-033-537A-1
Sequence 1, Application US/09033537A
Patent No. 5958083
GENERAL INFORMATION:

APPLICANT: Onishi, Masahiro
APPLICANT: Fich, Merete
APPLICANT: Toft, Annette Hanne
APPLICANT: Shlein, Martin
TITLE OF INVENTION: Prevention Of Back-Staining
TITLE OF INVENTION: In Stone Washing
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/033,537A
FILING DATE: 02-MAR-1998
CLASSIFICATION: 008
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0993/95
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: PCT/DK96/00364
FILING DATE: 03-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4492.204-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-033-537A-1

Query Match 5.7%; Score 118; DB 2; Length 551;
Best Local Similarity 18.7%; Pred. No. 0.011;
Matches 81; Conservative 47; Mismatches 119; Indels 186; Gaps 19;
QY 89 APAYNDGPTGLQGLDYA-----IASAAHGLRVIVVLTNDWKEFGMDQYDKWY-- 138
Db 1 APAPFGOLKVGQNLVQSGOAVQLVGMSSHGL-----QWYGNF 40
QY 139 -----LPYHONFYTDPTQOAYKNVW-----HLL 163
Db 41 VNKSSLOMRDNRNMGVNFRAAWYTAEDGYITDPSVKKNKKEAVQASIDGLYVITDWHIL 100
QY 164 NRVNSIT-----GVYTKNDPTIFAWELANEPRCVGSCTLTPTSGTCTOATVNW 211
Db 101 SDGNPTYKAQSKAFFQEMATLYGNTPNVI-YEIANEP-----NGNVSWADYKSY 149
QY 212 VDQMSAYVKSIDPNHMSVGDGE-----GFYIGS----- 239
Db 150 AEEVITAIRAIDPDGVVIVGSPWSDIHLAADNPVSHSNVMYALHEYSCTHGFQFLDRDI 209
QY 240 -----TQSGCWPNYNDPS-DCVDNALLRVKNIDFGTYHLYPNY---- 276
Db 210 TYAMNKGAAIEVTEWGTSDASGNGGPFPOSKEWID----FLNARKISWVNMSLADKYVTS 266
QY 277 -----W--GONADWGTQWIKDH1-----ANAAAIKPTILEEFGWQTPDR 314
Db 267 AALMPGASPTGGWTDQOLSESG-KWVRDQIRQATGGGSGNPTAPAPTNLIS-----ATAGN 321
QY 315 DSIVYQTW-----TOTVRTNGEAGNFWMLAGNVNGQYPN----YDGFNVY--PSST 361

Db 322 AOVSLTWNVASCATSVTVKRATTSGPYTNVATGVGTATSYTNTGLTNGTIIYVVSASNS 381
QY 362 ATVLASEALAI 374
Db 382 AGSSANSAQASAT 394
RESULT 11
US-09-134-078-63
; Sequence 63, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 956 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
US-09-134-078-63
Query Match 5.7%; Score 117.5; DB 4; Length 956;
Best Local Similarity 19.2%; Pred. No. 0.028;
Matches 76; Conservative 52; Mismatches 110; Indels 157; Gaps 19;
QY 3 AGFTVASCQGVFLNGLPY-RYGGTNNYLSYQSHADVDVLAQAQAMNLSVIRTWGFI 61
Db 48 SEGITSLAG-----NSLFSNAGDTSDFY-----NAETVDFL--AENWNSLIR-----IAM 92
QY 62 GSLDGSVPTIDGNKNGFYQYWDPTGAPAYNDGPTGLQG-LDYATASAAHGLRVI 120
Db 93 GVKEN-----WD---GGNGYIDSPQEQEAKIRKVIDAAIANGIYVII-- 131
QY 121 TNDWKEFGMDQYDKWYGLPYHDNFYTDPTQOAYKNWVHLLNRVNSTGTGYTK--NDP 178
Db 132 --DWH-----THEAEIYDQV-----DFFTRMADLYGDTPPNMYEINPEP 170
QY 179 TIFAWELANPRCVCSTGLPTSGTCTQATIVNWDQMSAYVKSIDPNHNVSVGDCGFYIG 238
Db 171 IQSNPV-----IKNYAEQVIAGIRSKOPDNLIIVGTSNYSQO 208

QY 239 STOGSGWPYND-----PSD-----GVDNALLRVKNIDFGTYHLYPNY 276
Db 209 VDVASADPISDTVAVYTLHFYAAFNPNDNLNRVAQTALDNNVALFV----- 254
QY 277 WQONADWGTOMIKDHIANAAAIKPTILEFEGWOTPDORSVYQVOTVOTVRTNGEAGNFW 336
Db 255 -----TEWGT-----ILNT-----GQCEPKEST-----NTW 276
QY 337 MLAGNVNGQYPNYDGFNVYYPSPSTATVLALEALA 371
Db 277 MAFLKEKIGISHANWSLSDKAPPETGVSVOAGQGV 311
RESULT 12
US-09-465-519-4
; Sequence 4, Application US/09465519
; Patent No. 6403355
; GENERAL INFORMATION:
; APPLICANT: HAGIHARA, Hiroshi
; APPLICANT: KITAYAMA, Kaori
; APPLICANT: HAYASHI, Yasuhiro
; APPLICANT: IGARASHI, Kazuaki
; APPLICANT: ENDO, Keiji
; APPLICANT: OZAKI, Katsuya
; TITLE OF INVENTION: NOVEL AMYLASES
; FILE REFERENCE: 2173-0118P
; CURRENT APPLICATION NUMBER: US/09/465,519
; CURRENT FILING DATE: 1999-12-16
; EARLIER APPLICATION NUMBER: 10-362487 JAPAN
; EARLIER FILING DATE: 1998-12-21
; EARLIER APPLICATION NUMBER: 10-362488 JAPAN
; EARLIER FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-465-519-4
Query Match 5.3%; Score 110.5; DB 4; Length 501;
Best Local Similarity 19.6%; Pred. No. 0.049;
Matches 74; Conservative 50; Mismatches 124; Indels 129; Gaps 21;
QY 20 YRYGCTN-----NYLSYQ---SHADVDDVLAQAQAMNLSVIRTWG--FIDIG 62
Db 196 FRFANTNMNRVDEENGNDYLLGSNIDFSHPVQDEL-----KDWGSMFTDEL 244
QY 63 SLDCSVPTIDGNKN--GFYQYWDPTGAPAYNDGPTGLQGLDYAIAASAAHGLRVI 120
Db 245 DLDGY--RLDAIKHIPFWYTSWVRHQRNADOD-----LFVV 280
QY 121 TNDWK-EFGMDQY---DKW-----YGLPYHDNFYTDPTQOAY--KNWVHLLNRVNSIT 170
Db 281 GEYKDDVGALEFLDEMNMWMSLFDVPLNYFYRASQCGSGYDMRNLIRGSLVEAHPMH 340
QY 171 GVTYKNDPTIFAWELANPRCVCSTGLPTSGTCTQATIVNWDQMSAYVKSIDPNHNVSV 230
Db 341 AVTFEON-----HDTOP--GESLESWADMFKPL-AYATIL----- 373
QY 231 GDEGFYIGSTQSGWP-----YNDPSDQVD-----NNALLRVKNIDFGTYHLYPNY 277
Db 374 -----TREGYPNVFYGDYIGIPNDNISAKMDIDELDARQNYATGTOHDYFDHW 424
QY 278 GONADWGTOMIKDHIANAAAIKPTILEE-----FGWOTPDORSVYQVOTVOTVRTNGEA-- 331
Db 425 DV-----VCMTRGSSSRPNNSGLATIMNSGPGSKWYVGRQNGQGTWTDLTGNNGASVT 479
QY 332 -----GW-NFWHLAGNVN 343
Db 480 INGDCWGEFFTGGSYS 496

```

DB 248 FCD-GNEWMR---KLACSYMSRVVWSEFG 272

RESULT 14
US-09-116-622-6
: Sequence 6, Application US/09116522
: Patent No. 6080567
: GENERAL INFORMATION:
: APPLICANT: Kofod, Lene V.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Christgau, Stephan
: APPLICANT: Heidt-Hansen, Hans P.
: APPLICANT: Heldt-Hansen, Hans P.
: APPLICANT: Dalboge, Henrik
: APPLICANT: Andersen, Lene N.
: APPLICANT: Si, Joan Q.
: APPLICANT: Jacobson, Tina
: APPLICANT: Munk, Niels
: APPLICANT: Mullertz, Anette
: TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
: TITLE OF INVENTION: ASPERGILLUS ACULEATUS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6080567o No. 6080567disk of No. 6080567th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: APPLICATION NUMBER: US/09/116.622
: FILING DATE: 16-July-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Agtis, Cheryl H.
: REGISTRATION NUMBER: 34,086
: REFERENCE/DOCKET NUMBER: 3954.224-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 296 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-116-622-6

Query Match 5.2%; Score 109; DB 3; Length 296;
Best Local Similarity 22.3%; Pred. No. 0.032;
Matches 60; Conservative 37; Mismatches 96; Indels 76; Gaps 16

QY 73 GNKNCFYQYW-DPSTGAPAYNDGPTGLOGLDYAIAS-----AAAGHLRVIVVLT 121
DB 47 GYSNGYIYSPFTDASGDVEYNSGAGGSYSVTSSASFVGGKGNPWSAH-----ITYS 102
QY 122 NDWKEFGMDQYDKWYG-----LPYH---DNFYTDPRTOQAYKKNWVHLLNRVNSITGV 172
DB 103 GSWTSTGNSYLSVYGWTTGTPVLEYIILEDYGEYNPWSAGTYKGSVYS-----DGS 154
QY 173 TYKNDPTIFANELANEPKRCVSGTLP-----TSCTCTQATIVNWDQMSAYVK--- 220
DB 155 TY----NIVTATRTNAPSIOGTATFTQYWSIROTKRVGGTGTVA-----NHFNAAKLG 204
QY 221 -SIDPNHMYVSGDEGYIGSTGSGHPYNDPSGVDNALLRVKNIDFGTYHLVYPNYWGQ 279
DB 205 MNLGTHNYQIVATEGY-----SSGSASITVAERAD--ILLR-----YMLY--LWHR 247

US-08-507-431-6
: Sequence 6, Application US/08507431
: Patent No. 5693518
: GENERAL INFORMATION:
: APPLICANT: Kofod, Lene V.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Christgau, Stephan
: APPLICANT: Heidt-Hansen, Hans P.
: APPLICANT: Dalboge, Henrik
: APPLICANT: Andersen, Lene N.
: APPLICANT: Si, Joan Q.
: APPLICANT: Jacobson, Tina
: APPLICANT: Munk, Niels
: APPLICANT: Mullertz, Anette
: TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
: TITLE OF INVENTION: ASPERGILLUS ACULEATUS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5693518o No. 5693518disk of No. 5693518th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: APPLICATION NUMBER: US/08/507.431
: FILING DATE: 15-FEB-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/002,800
: FILING DATE: 25-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Harrington, James J.
: REGISTRATION NUMBER: 38,711
: REFERENCE/DOCKET NUMBER: 3954.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 296 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-507-431-6

Query Match 5.2%; Score 109; DB 1; Length 296;
Best Local Similarity 22.3%; Pred. No. 0.032;
Matches 60; Conservative 37; Mismatches 96; Indels 76; Gaps 16;

QY 73 GNKNCFYQYW-DPSTGAPAYNDGPTGLOGLDYAIAS-----AAAGHLRVIVVLT 121
DB 47 GYSNGYIYSPFTDASGDVEYNSGAGGSYSVTSSASFVGGKGNPWSAH-----ITYS 102
QY 122 NDWKEFGMDQYDKWYG-----LPYH---DNFYTDPRTOQAYKKNWVHLLNRVNSITGV 172
DB 103 GSWTSTGNSYLSVYGWTTGTPVLEYIILEDYGEYNPWSAGTYKGSVYS-----DGS 154
QY 173 TYKNDPTIFANELANEPKRCVSGTLP-----TSCTCTQATIVNWDQMSAYVK--- 220
DB 155 TY----NIVTATRTNAPSIOGTATFTQYWSIROTKRVGGTGTVA-----NHFNAAKLG 204
QY 221 -SIDPNHMYVSGDEGYIGSTGSGHPYNDPSGVDNALLRVKNIDFGTYHLVYPNYWGQ 279
DB 205 MNLGTHNYQIVATEGY-----SSGSASITVAERAD--ILLR-----YMLY--LWHR 247
QY 280 NADWGTQWIKDHANAAAIKPKTILEEFG 308

```

```

DB 248 FCD-GNEWMR---KLACSYMSRVVVSEFG 272

RESULT 14
US-09-116-622-6
: Sequence 6, Application US/09116522
: Patent No. 6080567
: GENERAL INFORMATION:
: APPLICANT: Kofod, Lene V.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Christgau, Stephan
: APPLICANT: Heidt-Hansen, Hans P.
: APPLICANT: Heldt-Hansen, Hans P.
: APPLICANT: Dalboge, Henrik
: APPLICANT: Andersen, Lene N.
: APPLICANT: Si, Joan Q.
: APPLICANT: Jacobson, Tina
: APPLICANT: Munk, Niels
: APPLICANT: Mullertz, Anette
: TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
: TITLE OF INVENTION: ASPERGILLUS ACULEATUS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6080567o No. 6080567disk of No. 6080567th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: APPLICATION NUMBER: US/09/116.622
: FILING DATE: 16-July-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Agtis, Cheryl H.
: REGISTRATION NUMBER: 34,086
: REFERENCE/DOCKET NUMBER: 3954.224-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 296 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-116-622-6

Query Match 5.2%; Score 109; DB 3; Length 296;
Best Local Similarity 22.3%; Pred. No. 0.032;
Matches 60; Conservative 37; Mismatches 96; Indels 76; Gaps 16

QY 73 GNKNCFYQYW-DPSTGAPAYNDGPTGLOGLDYAIAS-----AAAGHLRVIVVLT 121
DB 47 GYSNGYIYFSTWDGASGDVEYSGAGGSYVTSWSSAFVGGKGNWPGSAHD-----ITYS 102
QY 122 NDWKEFGMDQYDKWYG-----LPYH---DNFYTDPRTOQAYKKNWVHLLNRVNSITGV 172
DB 103 GSWTSTGNSYLSVYGWTTGTPVLEYIILEDYGEYNPGSAGTYKGSVYS-----DGS 154
QY 173 TYKNDPTIFANELANEPKRCVSGTLP-----TSCTCTQATIVNWDQMSAYVK--- 220
DB 155 TY----NIVTATRTNAPSIOGTATFTQYWSIROTKRVGGTGTVA-----NHFNAAKLG 204
QY 221 -SIDPNHMYVSGDEGYIGSTGSGHPYNDPSGVDNALLRVKNIDFGTYHLVYPNYWGQ 279
DB 205 MNLGTHNYQIVATEGY-----SSGSASITVAERAD--ILLR-----YMLY--LWHR 247

US-08-507-431-6
: Sequence 6, Application US/08507431
: Patent No. 5693518
: GENERAL INFORMATION:
: APPLICANT: Kofod, Lene V.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Christgau, Stephan
: APPLICANT: Heidt-Hansen, Hans P.
: APPLICANT: Dalboge, Henrik
: APPLICANT: Andersen, Lene N.
: APPLICANT: Si, Joan Q.
: APPLICANT: Jacobson, Tina
: APPLICANT: Munk, Niels
: APPLICANT: Mullertz, Anette
: TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
: TITLE OF INVENTION: ASPERGILLUS ACULEATUS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5693518o No. 5693518disk of No. 5693518th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/507,431
: FILING DATE: 15-FEB-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/002,800
: FILING DATE: 25-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Harrington, James J.
: REGISTRATION NUMBER: 38,711
: REFERENCE/DOCKET NUMBER: 3954.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 296 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-507-431-6

Query Match 5.2%; Score 109; DB 1; Length 296;
Best Local Similarity 22.3%; Pred. No. 0.032;
Matches 60; Conservative 37; Mismatches 96; Indels 76; Gaps 16;

QY 73 GNKNCFYQYW-DPSTGAPAYNDGPTGLOGLDYAIAS-----AAAGHLRVIVVLT 121
DB 47 GYSNGYIYFSTWDGASGDVEYSGAGGSYVTSWSSAFVGGKGNWPGSAHD-----ITYS 102
QY 122 NDWKEFGMDQYDKWYG-----LPYH---DNFYTDPRTOQAYKKNWVHLLNRVNSITGV 172
DB 103 GSWTSTGNSYLSVYGWTTGTPVLEYIILEDYGEYNPGSAGTYKGSVYS-----DGS 154
QY 173 TYKNDPTIFANELANEPKRCVSGTLP-----TSCTCTQATIVNWDQMSAYVK--- 220
DB 155 TY----NIVTATRTNAPSIOGTATFTQYWSIROTKRVGGTGTVA-----NHFNAAKLG 204
QY 221 -SIDPNHMYVSGDEGYIGSTGSGHPYNDPSGVDNALLRVKNIDFGTYHLVYPNYWGQ 279
DB 205 MNLGTHNYQIVATEGY-----SSGSASITVAERAD--ILLR-----YMLY--LWHR 247
QY 280 NADWGQTQWKDHANAAAIKPKTILEEFG 308

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:55:11 : Search time 7.27371 Seconds
(without alignments)
776.473 Million cell updates/sec

Title: US-09-917-378-3

Perfect score: 2079

Sequence: 1 APAGFTVAGCGQVNLNGLPY.....YPPSSTATVLAASEALAISTG 375

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Scarched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications:AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	812	39.1	663	9	US-10-121-032-61
2	812	39.1	680	9	US-10-121-032-25
3	493	23.7	427	9	US-09-850-9828-2
4	179.5	8.6	562	10	US-09-981-9008-5
5	124.5	6.0	711	10	US-09-765-272-82
6	117.5	5.7	956	9	US-10-121-032-63
7	117	5.6	951	9	US-09-924-097-15
8	110.5	5.3	501	9	US-09-918-543-27
9	110.5	5.3	501	9	US-09-918-543-28
10	110.5	5.3	501	10	US-09-971-611-4
11	109	5.2	467	10	US-09-863-5478-1
12	104.5	5.0	501	9	US-09-918-543-25
13	104	5.0	1010	10	US-09-118-276-12
14	97	4.7	553	10	US-09-888-224-2
15	95	4.6	877	10	US-09-881-752A-28
16	93.5	4.5	485	9	US-09-918-543-2
17	93.5	4.5	485	10	US-09-769-864-1
18	93.5	4.5	485	10	US-09-769-864-7
19	93.5	4.5	485	10	US-09-854-346-2

20	93.5	4.5	485	10	US-09-902-188A-1
21	93	4.5	1430	10	US-09-740-274-6
22	92	4.4	514	10	US-09-769-864-3
23	92	4.4	514	10	US-09-902-188A-3
24	91.5	4.4	472	10	US-09-745-763-36
25	91.5	4.4	472	10	US-09-995-587A-11
26	91	4.4	515	9	US-09-918-543-6
27	91	4.4	515	10	US-09-854-346-6
28	89	4.3	925	9	US-09-924-097-14
29	88	4.2	386	10	US-09-739-861A-5
30	88	4.2	386	10	US-09-795-583-5
31	87	4.2	486	10	US-09-815-242-13455
32	86.5	4.2	672	10	US-09-815-242-10306
33	86.5	4.2	1375	10	US-09-740-274-4
34	85.5	4.1	654	9	US-09-905-291A-177
35	85.5	4.1	654	10	US-09-909-320-177
36	85.5	4.1	654	10	US-09-909-088B-177
37	85.5	4.1	654	12	US-10-052-586-26
38	85.5	4.1	1475	10	US-09-740-274-2
39	84.5	4.1	156	10	US-09-738-973-210
40	84.5	4.1	1073	10	US-09-815-242-12361
41	84.5	4.1	1147	10	US-09-815-242-5468
42	84.5	4.1	5795	10	US-09-815-242-12610
43	84	4.0	275	10	US-09-060-854B-4
44	84	4.0	357	10	US-09-837-235-15
45	84	4.0	516	10	US-09-986-676A-2

ALIGNMENTS

RESULT 1

US-10-121-032-61

: Sequence 61, Application US/10121032

: Patent No. US20020155550A1

: GENERAL INFORMATION:

: APPLICANT: Bylina, Edward J.

: TITLE OF INVENTION: GLYCOSIDASE ENZYMES

: NUMBER OF SEQUENCES: 72

: CORRESPONDENCE ADDRESSES:

: ADDRESSEE: Gray Cary Ware & Freidenrich LLP

: STREET: 4365 Executive Drive, Suite 1600

: CITY: San Diego

: STATE: CA

: COUNTRY: USA

: ZIP: 92121

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: Windows95

: SOFTWARE: FastSeq for Windows Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/10/121,032

: FILING DATE: 09-Apr-2002

: CLASSIFICATION: <Unknown>

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US/09/134,078

: FILING DATE: 13-AUG-1998

: APPLICATION NUMBER: 08/949,026

: FILING DATE: 10-OCT-1997

: APPLICATION NUMBER: 60/056,916

: FILING DATE: 06-DEC-1996

: ATTORNEY/AGENT INFORMATION:

: NAME: Halie, Lisa A.

: REGISTRATION NUMBER: 38,347

: REFERENCE/DOCKET NUMBER: 09010/024002

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 858/677-1456

: TELEFAX: 858/677-1465

: INFORMATION FOR SEQ ID NO: 61:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 663 amino acids

: TYPE: amino acid

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 61:
US-10-121-032-61

Query Match          39.1%; Score 812; DB 9; Length 663;
-Best Local Similarity 42.8%; Pred. No. 7.2e-62;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

Qy   5 FVTASGGQFVLNGLPYRGYGTGNTYYLSQSASHADVDDVLAKAAQNLSVIRTGWGIDIGSL 64
    || :||| :|| :||||| :|| :||| :|| :||| :||| :||| :||| :||| :||| :|||
Db   17 FVKVENCKFALNGKREFRFGISNNYYMHYKXNGMIDSVLESARDMGIKVLRWGF----L 71

Qy   65 DGSVPTTIDGNKNGYFYQWDPSTCAPAYNDG----PTGLQGLDYAIAASAAHGLRVIVVL 120
    || :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   72 DGESYCRDKN-----TYMHPEPGVFCVPECFNAOQSGFERKLDTYVAKAKELGIKLVLV 125

Qy   121 TNDMKERGGMDYDKWYGLPYHDNFYDPRTOQAQYKNWNHILLNKRNSITGVTVKNDPTI 180
    || :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   126 VNNWDDFGGMNQYVRFGCTHHDDFYDEKIKEEYKYVSFLVNHVNTYTTCVPREEPTI 185

Qy   181 PAWELANEPVCVGSGTLPTSCCTCTQATIVNVVDQMSAYKSIDPNHMVSVSGDEGF---YI 237
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   186 MAWELANEPRC-----ETDKSGN-----TLVEWVKEMSSYIKSLDPNHILVAVGDGEGFSNYE 237

Qy   238 GSTOGSG--WPYNDDPSGDVNNALLRVKNKTDFCTYHLYPNWYGN-----ADMCTOWTKD 290
    || :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   238 GFKPYPGEAEWAYNGWS-GVDWKLLSLETVDFTGTHLYPSHWGVSPEHYAQWGAKWIIE 296

Qy   291 HIANAAAIGRPRTILEEFG--WOTP-DRDSVYQTWTQTVRTNAGEAGNFMWMLAGNVNGOP- 346
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   297 HKIKAKEIGRPVLEEEYGIPKSAPVNRATYRLWNLVDLYDLGGDGAMFWMLAGICEGSDR 356

Qy   347 -----YPNYDGFNNYYPSSSTATVLASE--ALAISTG 375
    ||:|||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   357 DERGYPDYDGRVINDDSPAELIREYAKLFNTG 391


RESULT 2
US-10-121-032-25
; Sequence 25 Application US/10121032
; Patent No. US20020155550A1
; GENERAL INFORMATION:
; APPLICANT: Hyline, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10121.032
; FILING DATE: 09-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/134.078
; FILING DATE: 13-AUG-1998
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
```

```

: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 858/677-1456
:   TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 680 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-10-121-032-25

Query Match          39.1%; Score 812; DB 9; Length 680;
Best Local Similarity 42.8%; Pred. No. 7.5e-62;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

QY    5 FVTASGCGFVLNGLPYRYGGTNNYYLSYSQSHADVDDYLAKAQAANLNSVIRTWGFIDIGSL 64
      || |::||| :|:||||| ::|: |::| |::| |::| |::| |::| |::| |::| |::| |
Db    34 FVKVENGKEALNGEKFRIFGSNYYMHYKSGMIDSVLESARDNGCIKVLRTWCF-----L 88

QY    65 DGSVPTIDGNKNRGFYQVWPSTCAPAYNDG----PTGLOGLDYAIASAAGILRVIVLI 120
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db    89 DGESYCRDKN-----TYMHPEPCVFCVPEGISNAOSGFERLDYTVAKAKELGIKLIVL 142

QY    121 TNDWKKEFGMDQDYDKWYGLPYHDNFYTDPRTOQAYKWNHNLNRVWSITGVITYKNDIPTI 180
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db    143 VNNWDFEGGMNQYYVRVFEGTHDDFDIKEYEKKYVSFLVNHHVNTYTCVPYREPTI 202

QY    181 FAWELANEPCRCVSGTLPTSGTGCTQATIVNWVDQMSAVKSIDPNHMVSYDCGEF---YI 237
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db    203 MAHELANEPRC-----ETDKSGN----TLVEVKEMSSYIKSLDPNHLVAVGDGEFFSNYE 254

QY    238 GSTOGSG---WPYNDRPSGDVDNALLRKVIDFGTYHLYPNYWGON----ADMGTOWIKD 290
      | |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db    255 GFKPYGGEAEWAYNGWS-GVDMKKLLSIETVDECTFHPYSHMGVSPENYAQMCAKWIED 313

QY    291 HIANAAIGKPTILEEFC--WOTP-DROSVYOTWTQVTRTNGEAGHWFMWLACNVGNCP- 346
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db    314 HKIAKEIGKPVLEEYGIKPSAPVNTAIRUMDLIVLYDLGGDGAMFWMLAGIGEGR 373

QY    347 -----YPNYDGFNYYYSPSSTATVLASE-ALAISTG 375
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db    374 DERCYYPDYDGRFVNDSDSPEALI REYAKLFNTG 408


RESULT 3
US-09-850-982B-2
: Sequence 2, Application US/09850982B
: Patent No. US20020166145A1
: GENERAL INFORMATION:
: APPLICANT: Nestec S.A.
: TITLE OF INVENTION: COFFEE MANNAHAVE
: FILE REFERENCE: 88265-4025
: CURRENT APPLICATION NUMBER: US/09/850,982B
: CURRENT FILING DATE: 2001-05-08
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 427
: TYPE: PRT
: ORGANISM: Coffea arabica
US-09-850-982B-2

Query Match          23.7%; Score 493; DB 9; Length 427;
Best Local Similarity 32.1%; Pred. No. 8.2e-35;
Matches 125; Conservative 62; Mismatches 137; Indels 66; Gaps 16;

QY    5 FVTASGCGFVLNGLPYRYGGTNNYYLSY-----QSHADVDDYLAKAQAMNLSVIRTWGFI 59
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db    38 FIQTRGTRFVLGGYPFFFNFGFSNMWHMVAAEFSERHKISNWFREAAATGLTVCRTWAFS 97

```

```
: : REFERENCE/DOCKET NUMBER: 09010/024002
: : TELECOMMUNICATION INFORMATION:
: : TELEPHONE: 858/677-1456
: : TELEFAX: 858/677-1465
: : INFORMATION FOR SEQ ID NO: 25:
: : SEQUENCE CHARACTERISTICS:
: : LENGTH: 680 amino acids
: : TYPE: amino acid
: : TOPOLOGY: linear
: : MOLECULE TYPE: protein
: : FRAGMENT TYPE: internal
: : SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-10-121-032-25

Query Match          39.1%; Score 812; DB 9; Length 680;
Best Local Similarity 42.8%; Pred. No. 7.5e-62;
Matches 169; Conservative 61; Mismatches 121; Indels 44; Gaps 13;

Qy   5 FVTASGCGFYVLNGLPYRGYGGTNNYYLSYSQSADVDDYLAKAQAANLNSVIRTWGFIDIGSL 64
      || |::||| ::|:||||| |:|: |::| |::| |::| |::| |::| |::| |::| |::| |
Db   34 FVKVENGKFAI NGKEFRFGISNNNYMHYKSGMIDSLESARDNGCI KVLRIWCF-----L 88

Qy   65 DGSVPTIDGNKNNGFYQVWPDPSTCAPAYNDG----PTGLGCLDYAIASAAGILRVIVLI 120
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db   89 DGESYCRDKN-----TYMHPFCVGVPEGISNAOSGFERLDYTVAKAKELGIKLVI L 142

Qy   121 TNDWKFFGGMDQDKYGLYPHYDNFYDTPTQQAYKWVNHLNRVWSITCVITYKNIDPPI 180
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db   143 VNNWDFFGGMNQYVRVFEGTHHDFRDEKI KEEYKKYVSFLVNHVNTYTCTVPYREPTI 202

Qy   181 PAWEIANEPRCVSGSGLTPTSCTCTCAATIVNVVDQMSAVKSIDPNHMVSVDCGEF---YI 237
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db   203 MAWEIANEPRC-----ETDKSCN----TLVEVKEMSSYIKSLDPNHLVAVGDGEFFSNYE 254

Qy   238 GSTOGSG---WPYNDRPSGDVDNNALLRKNI DFCTYHL YPNYWGON----ADMGTOWIKD 290
      | |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db   255 GFKPYGGAEWAYNGWS-GVDMKKLLSI ETGTFPHLYPSHWGVS PENYAQMCAKWI ED 313

Qy   291 HIAANAIGKPTILEEFC--WOTP-DROSYVOYTQTVRTNGEAGHNFWMLAGNVCNP- 346
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db   314 HKIAKEIGKPVLEEYGI PKSA PNRTAIRUMDLIVLYDLGGDGAMPFMLAGIGEGR 373

Qy   347 -----YPNYDGFNYYYSPSATVLASE-A LAISTG 375
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db   374 DERCYYPDYDGRFVNDDSDPEAE LI REYAKLFNTG 408


RESULT 3
US-09-850-982B-2
: : Sequence 2, Application US/09850982B
: : Patent No. US20020166145A1
: : GENERAL INFORMATION:
: : APPLICANT: Nestec S.A.
: : TITLE OF INVENTION: COFFEE MANNA NASE
: : FILE REFERENCE: 88265-4025
: : CURRENT APPLICATION NUMBER: US/09/850,982B
: : CURRENT FILING DATE: 2001-05-08
: : NUMBER OF SEQ ID NOS: 12
: : SOFTWARE: PatentIn version 3.1
: : SEQ ID NO 2
: : LENGTH: 427
: : TYPE: PRT
: : ORGANISM: Coffea arabica
US-09-850-982B-2

Query Match          23.7%; Score 493; DB 9; Length 427;
Best Local Similarity 32.1%; Pred. No. 8.2e-35;
Matches 125; Conservative 62; Mismatches 137; Indels 66; Gaps 16;

Qy   5 FVTASGCGFYVLNGLPYRGYGGTNNYYLSY-----QSHADVDDLAKAQAAAMLNSVIRTWGF 59
      || |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
Db   38 FIQTRGTRFVLGGYPFPFNFGFSNMWHMVAAEFSERHKISNWFEAAAATGLTCVWTWAFS 97
```


Db 589 YKALRAKHPKWLIIYGSETSSATRTKSYRPERELKHSNCPERNYQSDYGNDRVCGWGT 648
Qy 281 AOWGTOWIKDHIANAAA-----IGKPTILEEFQWQTPDRDSVYQVWTQVTRTNGEAG 332
Db 649 ATASWTFDRDAGYAGQFIWTCDTDICEPT-----PWHNQNTTPVKSSYFEGIVDTAGIPK 703
Qy 333 WNFWM 337
Db 704 HIFYL 708

RESULT 6
US-10-121-032-63
; Sequence 63, Application US/10121032
; Patent No. US2002015550A1
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/121,032
; FILING DATE: 09-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 956 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 63:
US-10-121-032-63

Query Match 5.7%; Score 117.5; DB 9; Length 956;
Best Local Similarity 19.2%; Pred. No. 0.031;
Matches 76; Conservative 52; Mismatches 110; Indels 157; Gaps 19;

Qy 3 AGFVTASGGQFVLNGLPY-RYGGTNNYLSYQSHADVDVLAQAAAMNLSVIRTWGFI 61
Db 48 SGEITSLAG-----NSLFWNAGDTSDFY-----NATVDFL--AEWNSSLIR-----IAM 92
Qy 62 GSLDGSVPTIDGNKNGFYQYWDPTSGAPYNDGPTGLQGLD-LDYAIAASAAHGLRIVVYL 120
Db 93 GYKEN-----WD--GGNGYIDSPQEQEAKIRKVIDAALANGIYVII-- 131
Qy 121 TNDWKEFGMDQYDKWGLPYHDNFYDPTQQAAYKNWNHLLNRVNSITGVTYK--NDP 178

Db 132 --DMH-----THEAELYTDEAV-----DFFTRMADLYCDTNPVNYEIVNEP 170
Qy 179 TIFAWELANEPKRCVSGTLPSTCTCTQATIVNVDMSAYVKSIDPNIMVSGDEGFIYIC 238
Db 171 IYQSWPV-----IKNYAEQVIAGIRSKDPDNLIIVGTSNYSQO 208
Qy 239 STOGSGMPYND-----PSD-----CVDNNALLRVKNIDFGTYHLYPNY 276
Db 209 VDVASADPISDTNWAYTLHFYAAFNPNDNLNRNVAQTALDNNVALFV----- 254
Qy 277 WGNADKWGTQWKDHIANAAAIGKPTILEEFQWQTPDRDSVYQVWTQVTRTNGEAGNFW 336
Db 255 ----TEWGT-----ILNT-----GQEPDKEST-----NTW 276
Qy 337 MLAGNVNGQYPNVDGFNVYVSPSTATVLASEALA 371
Db 277 MAFLKEKGISHANWSLSKAPFETGVSVOAGOGVS 311

RESULT 7
US-09-924-097-15
; Sequence 15, Application US/09924097
; Patent No. US20020156240A1
; GENERAL INFORMATION:
; APPLICANT: TOMONO, Jun
; APPLICANT: NOMURA, Yoshiko
; APPLICANT: SAGAWA, Hiroaki
; APPLICANT: SAKAI, Takeshi
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: ALPHA-AGARASE AND PROCESS FOR PRODUCING THE SAME
; FILE REFERENCE: TOMONO-1
; CURRENT APPLICATION NUMBER: US/09/924,097
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: JP00/00966
; PRIOR FILING DATE: 2000-02-21
; PRIOR APPLICATION NUMBER: 11-44890
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: 11-198852
; PRIOR FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 951
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of agarase 4-3
US-09-924-097-15

Query Match 5.6%; Score 117; DB 9; Length 951;
Best Local Similarity 20.1%; Pred. No. 0.033;
Matches 74; Conservative 45; Mismatches 113; Indels 136; Gaps 21;

Qy 4 GFVTASGG---QFVLNGLPYRYGCTNNYIL-----SYQSHADVDVLA 43
Db 209 GFVAAGATNVSMTNG---DFGDYNNIYLEAGTYRSFITVATGSDGSGYCARIDLDGSPA 265
Qy 44 KAAMNLSVIRTWGFIIDIGSLDGSVPTIDGNKNGFYQYWDPTSGAPYNDGPTGLQGLD 103
Db 266 -----SMGYFD--STGG-----WETPEEVELY-----GGD 288
Qy 104 YATASAAHGLRIVVLTNDWKEFGMDQYDKWGLPYHDNFYDPTQQAAYK-----NW 158
Db 289 FVVTSGTFLRVEAIGGSDM-----QWSG-----DNVRLTKVGDASVKPSPLYNP 334
Qy 159 VNHLLNRVNS-ITGVITYKNDPTIFAWELANEPKRCVSGTLPSTCTCTQATIVNVDMSA 217
Db 335 KDHTVTEIEGPVVGLEFLKKPV-----QVPTANRLKSDV--WYTY--- 373
Qy 218 YVKSIDPNIMVSGDEGFIYIGTQSGWPNVNDPSDGVNNALLRVKNIDFGTYHLYPNY 277
Db 374 -----PONNELQGFQNF--GAT-GSFWGHPPPEENFYDOTTI-----IDW--TOLVONYQ 417


```
Best Local Similarity 19.6%; Pred. No. 0.054;
Matches 74; Conservative 50; Mismatches 124; Indels 129; Gaps 21;

QY 20 YRYGGTN-----NYILSYQ---SHADVDDVLAKAAMNLSVIRTWG--FIDIG 62
DB 196 FRFANTNMNRVDEENGNYDYLLGSNIDFSHPEVDEL-----KDWGSWFTDEL 244
QY 63 SLDGSVPTIDGKNK--GFYFQYWDSPSTGAPAYNDGPTGLOGLDYAIASAAHCLRVIVVL 120
DB 245 DLDGY--RLDAIKH1PFWYTSWVRHORNEADQD-----LFVV 280
QY 121 TNDWK-EFGGMDQY---DKW---YGLPYHDNFTDPTQOAY--KNWNHLLNRVNSIT 170
DB 281 GEYKDDVGALFYLDEMNMWMSLFDVPLNYNFRASQCGSYDMRNILRGSLVEAHPIH 340
QY 171 GVTYKNDPTIFAWELANPRCVSGTLPSTGCTCTOATIVNWDQMSAYVKSIDPNHMVS 230
DB 341 AVTFVDN-----HDTQP--GESLESWADWFKPL-AYATIL----- 373
QY 231 GDEGFYIGTQSGWP-----YNDPSCVD-----NNALLRVKNIDFGTYHLYPNYW 277
DB 374 -----TREGGYPNVFYGDYIGIPNDNISAKKMDIDELLDARQNYAYGTOHDFDHW 424
QY 278 GQADWGTOWIKDHIANAALCKPTILEE-----FGQHTPDRDSVYQTWTQVTRTNGEA-- 331
DB 425 DV-----VGTWREGSSSRPNSGLATIMSNPGGSKMYYVGRONAGOTWTDLTGNCASVT 479
QY 332 ----GW-NFWMLAGNVN 343
DB 480 INGDCWGEFFTNGGSVS 496

RESULT 11
US-09-863-547B-1
; Sequence 1, Application US/09863547B
; Patent No. US20020128166A1
; GENERAL INFORMATION:
; APPLICANT: Henkel KGaA
; APPLICANT: Hermanus, Lentus B. M.
; APPLICANT: Van Beckhoven, Rudolf F. W. C.
; APPLICANT: Maurer, Karl-Helmut
; APPLICANT: Kottwitz, Beatrix
; APPLICANT: Weiss, Albrecht
; APPLICANT: Van Solingen, Pieter
; TITLE OF INVENTION: Detergents Comprising Cellulases
; FILE REFERENCE: H 1920 A
; CURRENT APPLICATION NUMBER: US/09/863,547B
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 08/945,574
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: PCT/EP96/01755
; PRIOR FILING DATE: 1996-04-26
; PRIOR APPLICATION NUMBER: US 08/614,115
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: EP 95201115.3
; PRIOR FILING DATE: 1995-04-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Bacillus sp. CBS 670.93
US-09-863-547B-1

Query Match 5.2%; Score 109; DB 10; Length 467;
Best Local Similarity 22.3%; Pred. No. 0.066;
Matches 59; Conservative 31; Mismatches 96; Indels 78; Gaps 16;

QY 87 TGAPAYNDGPTGLOGLDYAIAASAAHCLRVIVVLNWDKMEFGMDQYDKWGLPYHDNPFY 146
DB 93 TSSGGYIDDPVSKVKETVEAIDLGIYVII-----DWHILS-----DN-- 132
QY 147 TDPRTQOAYKNWNHLLNRVNSITGVTYKNDPTIFAWELANPRCVSGTLPSTGCTQOA 206

Db 133 -DP---NIYKEAKDFDEMSLEYG-DYPN-----VIVEIANEPN--GSD----- 170
QY 207 TIVNWDQMSAYVKS-----DPNIMVSGDEGFYIGTQSGWPNYPNPPSGVDNNAL 259
DB 171 --VTMDNQIKPYAEEVIPVIRONDPNNVIV-----CTGTWSDOVHHAADNQ-- 215
QY 260 LRVKNIDFCTYHLYPNYWGONADWGTOWIKDHIANAALCKPTILFEEFGWQTPDRD-SVY 318
DB 216 LADPNVY-AFHYAGTGHON-----LRDQVDYALDOGAALFVSEWGTSAATGDCGVF 267
QY 319 ----QTWTQTVRTNGEAGWFWML 338
DB 268 LDEAQVWIDFM-DERNLSNANWSL 290

RESULT 12
US-09-918-543-25
; Sequence 25, Application US/09918543
; Patent No. US20020155574A1
; GENERAL INFORMATION:
; APPLICANT: NO. US20020155574A1ozymes A/S
; APPLICANT: Thisted, Thomas
; APPLICANT: Kjaerulf, Soren
; APPLICANT: Andersen, Carsten
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Alpha-amylase mutants with altered properties
; FILE REFERENCE: 10062.200-US
; CURRENT APPLICATION NUMBER: US/09/918,543
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-918-543-25

Query Match 5.0%; Score 104.5; DB 9; Length 501;
Best Local Similarity 19.1%; Pred. No. 0.17;
Matches 72; Conservative 52; Mismatches 124; Indels 129; Gaps 20;

QY 20 YRYGGTN-----NYILSYQ---SHADVDDVLAKAAMNLSVIRTWG--FIDIG 62
DB 196 FRFANTNMNRVDEENGNYDYLLGSNIDFSHPEVDEL-----KDWGSWFTDEL 244
QY 63 SLDGSVPTIDGKNK--GFYFQYWDSPSTGAPAYNDGPTGLOGLDYAIASAAHCLRVIVVL 120
DB 245 DLDGY--RLDAIKH1PFWYTSWVRHORSEADQD-----LFVV 280
QY 121 TNDWK-EFGGMDQY---DKW---YGLPYHDNFTDPTQOAY--KNWNHLLNRVNSIT 170
DB 281 GEYKDDVGALFYLDEMNMWMSLFDVPLNYNFRASQCGSYDMRNILRGSLVEAHPIH 340
QY 171 GVTYKNDPTIFAWELANPRCVSGTLPSTGCTCTOATIVNWDQMSAYVKSIDPNHMVS 230
DB 341 AVTFVDN-----HDTQP--GESLESWADWFKPL-AYATIL----- 373
QY 231 GDEGFYIGTQSGWP-----YNDPSCVD-----NNALLRVKNIDFGTYHLYPNYW 277
DB 374 -----TREGGYPNVFYGDYIGIPNDNISAKKMDIDELLDARQNYAYGTOHDFDHW 424
QY 278 GQADWGTOWIKDHIANAALCKPTILEE-----FGQHTPDRDSVYQTWTQVTRTNGEA-- 326
DB 425 DI-----VGTWREGSSSRPNSGLATIMSNPGGSKMYYVGRONAGOTWTDLTGNCIASVT 479
QY 327 TNGEAGWFWMLAGNVN 343
DB 480 INGDCWGEFFTNGGSVS 496

RESULT 13
US-09-118-276-12
```


Sequence 12, Application US/09118276
Patent No. US20010011381A1
GENERAL INFORMATION:
APPLICANT: BABYCHUK, ELENA;
APPLICANT: KUSHNIR, SERGEI;
APPLICANT: DE BLOCK, MARC;
APPLICANT: INZE, DIRK
TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
CELL DEATH IN EUKARYOTIC CELLS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: SIXBEY, FRIEDMAN, LEEDOM, & FERGUSON
STREET: 8180 GREENSBORO DRIVE, SUITE 800
CITY: MCLEAN,
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22102
COMPUTER READABLE FORM:
MEDIUM TYPE: 3-1/2" DISKETTE
COMPUTER: IBM-COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/118,276
FILING DATE: 17-JUL-1998
PRIOR APPLICATION DATA: NONE
ATTORNEY/AGENT INFORMATION:
NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
REGISTRATION NUMBER: 31,196; 43,077
REFERENCE/DOCKET NUMBER: 6201-0003
TELEPHONE: (703) 790-9110
TELEFAX: (703) 883-0370
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1010 RESIDUES
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-09-118-276-12

Query Match 5.0%; Score 104; DB 10; Length 1010;
Best Local Similarity 19.3%; Pred. No. 0.47;
Matches 70; Conservative 49; Mismatches 127; Indels 116; Gaps 16;
QY 6 VTASGGQFVLNGLPYRYGCTNNYLSYQSHADVDVLAQAQANLNSVIRTWGFDIGSLD 65
DB 681 VAVKGEQFLINHPPF-----YFTGGRHEDAD--LRGKGFNVLMVHDHALMD----- 726
QY 66 GSVPTIDGNKNGFYQYWDPTGAPAYNDG--PTGLQGLDYAIASAAAHGLRVIVLTND 123
DB 727 -----W-----ICANSYRTSHYPYAEMLDW----ADEHG---IVVIDE 758
QY 124 WKEFGGMDQYKYGLPYH-----DNFYTD-----RTQQAQKNNVHLLNRVNSITGVTY 174
DB 759 TAAVG-----FNLSLGIGFEAGNPKELYSEAVNGETQQAHLQAIKELIAR-----D 806
QY 175 KNDPTIFAMELANEPRCVSGTLPTSCCTQATIVNVDQMSAYVKSIDPNHNVSVGDEG 234
DB 807 KNIHPSVVMMSIANEPDTRQAGREYFAPLAEAT-----RKLDPTRTITCVNM 854
QY 235 FYIGSTQSGWPNDSGDVNNALLRVKNIDFGTYHLYPNYWGQADWGTQ---WIKDH 291
DB 855 FCAHTP-----DTISDLFDVLCLNR-----YGVYVQSCDLETAEKVLEKEL 896
QY 292 IANAAAIKPTILEEFG-----NQTPDRDSVYQTWGTQVTRTNGEAGW 333
DB 897 LAWQEKLHPIIITIEYGVDTLAGLHSMYTDMMSEYOACALMDMYHRVDFRVSAAVGEQW 956
QY 334 NF 335
DB 957 NF 958

RESULT 14
US-09-888-224-2
Sequence 2, Application US/09888224
Patent No. US20020120118A1
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/888,224
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/430,669
FILING DATE: 28-OCT-1999
APPLICATION NUMBER: US/08/651,572
FILING DATE: 22-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS: Unknown
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-888-224-2

Query Match 4.7%; Score 97; DB 10; Length 553;
Best Local Similarity 18.4%; Pred. No. 0.87;
Matches 83; Conservative 49; Mismatches 142; Indels 178; Gaps 24;
QY 25 TNNYLSYQSHADVDVLAQAQANLNSVIRTWGFDIGSLDGSVPT-IDGNKNGFYQYW 83
DB 24 TPNVYVHGLSNRMEDMLQIKSLGFNAIRL-PECTOSVXKPGTMTAIDYAKN----- 75
QY 84 DPSTGAPAYNDGPTGLQGLDYA-----IASAAAHGLRVIVLTNDNKEFGGMDQYDKWY 137
DB 76 -----PDLOGLSVQIMEKLIKAGDLGIFVL----- 102
QY 138 GLPYHD---NF-----YTDPTQQAQYKN-WVN--HLLNRVNSITGVTYKNDPTIFAMELA 186
DB 103 -LDYHRIQCNFIEPLWYTFDSFSEQDYINTWVEVAQREKYNVNVIGADLKNKP-----HSS 157
QY 187 NEPRCVSGTLPSTGCTQATIVNW---VDMSAYVKSIDPNHNVSVGDSGFYIGSTOG- 242
DB 158 PAPAAYTDGSGATWGMGNAT--DWNLAERICRAILEVAPQWYFVEGTQFTTPEIDGR 215
QY 243 -----SGW-----PYNDPSDGVNHNALLRVK 263
DB 216 YKWHNNAWCGNLMGVRYKYPVNLPRDKVYVSPQYGVSEVYDQYFDPGEGFPDN----- 269
QY 264 NIDFGTYHLYPNYWGQADWGTQWIKDHIANAAAIKPTILEEFG----- 308
DB 270 -----LPEIWIYHH--FG--YVK-----LDLGYPVYVIGEGGKYGHGGDRDVTWQ 310

Search completed: November 13, 2002, 12:08:33
Job time : 10.2737 secs

Oy	309	-----WQTPDR--DSVYOTWTQTVRNGEAG-----W-NFW-----MLAGNV	342
		: : : : :	:
Db	311	NKIIDMIQNKFCDFFYSWNP---NSGDTGGILKDDTTIWEDKYNNLRLMDSCSGNA	367
Oy	343	NGQPYPNYDGNFVVPYSSSTATVLASEALAI	374
Db	368	TATSVPPTTITTSPTPTTTTTSTPTTTQT	399
		: : : : :	:

RESULT 15

```

US-09-881-752A-28
: Sequence 28, Application US/09881752A
: Patent No. US20020115078A1
: GENERAL INFORMATION:
: APPLICANT: Kleenthous, Harold
: APPLICANT: Al-Garawi, Amal
: APPLICANT: Miller, Charles
: APPLICANT: Tomb, Jean-François
: APPLICANT: Omenen, Raymond P.
: TITLE OF INVENTION: Identification of Polynucleotides
: TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the
: TITLE OF INVENTION: Genome
: FILE REFERENCE: 06132/041002
: CURRENT APPLICATION NUMBER: US/09/881.752A
: CURRENT FILING DATE: 2001-06-15
: PRIOR APPLICATION NUMBER: US 08/833,457
: PRIOR FILING DATE: 1997-04-01
: NUMBER OF SEQ ID NOS: 370
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 28
: LENGTH: 877
: TYPE: prt
: ORGANISM: Helicobacter pylori
US-09-881-752A-28

```

Query Match	4.68;	Score 95;	DB 10;	Length 877;
Best Local Similarity	20.9%;	Pred. No. 2.3;		
Mat.hes	89;	Conservative	43;	Mismatches 104; Indels 190; Gaps 23;
Qy	5	FVTASGGQF-VLNGLPYRYGGTNNYLSYQSHADYDDVLAKAQAMNLSVIRTWGFI	DIGS 63	
Db	462	FLINSGVNFKVTHPISEYDGNVFEYGMIOQ-----NLSVF-----	496	
Qy	64	LDGSVPTIDGNKNGFYQYWDPSGAPAYND--GPTGLQGLDYLAIASAAHGLRVIV	VLT 121	
Db	497	-----SLDGKNGYYKNNIDP-----NDPNGP-----	519	
Qy	122	NDWKEFGMDQYDKWYGLPYHDHNFYDTPRTOQAYKNWNHLLNRVNSTGTYTKNDP--	178	
Db	520	-----GLPYR-HYYTDOSSQYPQN-----LNTPNPL-----YRNMPPONS	552	
Qy	179	-----TIFAWELAN-----EPRCVSGSL-----PTSGTCTQATIVNWVDQM	215	
Db	553	HAIGNIIIGGFMOANYNILSNVIVGACTRYDIYLLDKNGRTHVTSGFSPSATVLY--	607	
Qy	216	SAYVKSIDPNHMVSGDGEFYI-----GSTOGSGWPNDDPSGDVNNALLRYK-----	NID 266	
Db	608	-----NPIESIGLKVSAYVTKALPGDGVLMRDPITYIYORNLRPAIQNWEPND	658	
Qy	267	FGTYHL-----YPNWQGNADGCTOWIKOHIANAAAI	GKPTILEEFQWOTPD 313	
Db	659	FSNSYFNVRGAAYQVYINNFINSYGQDTS-----KNGGGNATA-----KNMSGNLPE	705	
Qy	314	RDSYVOTQTQVTRTNGEAGNFWMLACGNVNGQYPNYDG-FNV--YYP	SSSTATVLA-SEA 369	
Db	706	TINYG-----YEVGNVR---YKNFLGTFSVARSWMTARGHLLADTYA	746	
Qy	370	LAISTG	375	
Db	747	LAATG	752	

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:49:26 ; Search time 13.829 Seconds
(without alignments)
1483.879 Million cell updates/sec

Title: US-09-917-378-4

Perfect score: 829

Sequence: 1 VSGGVKVOYKNDSAPGDNQ.....TKVTYVYNGRLVMCTEPTSGT 154

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

```

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374	45.1	551	AAW18790	Corrected Bacillus
2	373.5	45.1	700	AA13227	Novel endoglucanase
3	373	45.0	782	AA15625	Cellulase AP-1. A
4	369	44.5	1352	AA63962	Amino acid sequenc
5	368	44.4	677	AAU98063	Bacillus subtilis
6	360	43.4	499	AA12122	NK-1 cellulase. B
7	357.5	43.1	476	AA154123	A mannanase-linker
8	357.5	43.1	493	AA128850	Pectate lyase-link
9	357.5	43.1	493	AA13218	Pectate lyase CBD
10	357	43.1	1350	AA63963	Amino acid sequenc

11	356.5	43.0	167	17	AA95080	Cellulose binding
12	355.5	42.9	531	18	AAW15238	Scaffoldin protein
13	355.5	42.9	1853	19	AAW43108	C. thermocellum ce
14	349	42.1	616	20	AAW13494	Truncated cellulase
15	349	42.1	616	23	AAE16325	Active cellulase h
16	348	42.0	1751	20	AAW13493	Truncated cellulase
17	348	42.0	1751	23	AAE16324	Active cellulase p
18	340	41.0	1426	20	AAW13492	Truncated cellulase
19	340	41.0	1426	23	AAE16323	Active cellulase p
20	226	27.3	162	20	AAW90077	C. cellulovorans C
21	225	27.1	162	15	AAE63634	Cellulose binding
22	225	27.1	163	22	AAE05745	Clostridium cellu
23	225	27.1	328	22	AAE05749	Chimeric S peptide
24	225	27.1	341	22	AAE05747	Clostridium cellu
25	225	27.1	382	20	AAW39952	Gaussia luciferase
26	225	27.1	382	23	AAE13383	Gaussia species CB
27	225	27.1	428	22	AAE05748	Clostridium cellu
28	224	27.0	156	20	AAW90080	C. cellulovorans C
29	224	27.0	190	22	AAE05746	Clostridium cellu
30	218	26.3	154	20	AAW90081	C. cellulovorans C
31	127	15.3	531	16	AAW1503	60 kD endoglucanase
32	127	15.3	532	12	AAW13229	Endoglucanase enco
33	120.5	14.5	256	22	AAE81128	C17E2 OsPA constru
34	120.5	14.5	256	23	AAU97869	E. coli codon opti
35	101	12.2	646	22	AAE50866	Bacillus lichenifo
36	90.5	10.9	986	21	AAW40440	Human brain-deri
37	90.5	10.9	1131	23	AAE23672	Human secretin rec
38	90.5	10.9	1346	22	AAU04567	Human G-protein co
39	90.5	10.9	1346	22	AAU04581	Human G-protein co
40	90.5	10.9	1371	22	AAU04570	Human G-protein co
41	85.5	10.3	665	15	AAW47189	Cellulase AEC3. A
42	85	10.3	2123	22	AAE00701	Moraxella catarrha
43	84.5	10.2	475	22	AAE92298	C glutamicum prote
44	84	10.1	434	19	AAW50008	Human hyaluronidas
45	84	10.1	1992	17	AAW04505	Moraxella 200 kDa

ALIGNMENTS

RESULT 1
AAW18790
ID AAW18790 standard; protein; 551 AA.
AC AAW18790;
XX
DT 18-NOV-1997 (first entry)
XX
DE Corrected Bacillus lautus (NCIMR 40250) endoglucanase Endo 3A.
XX
KW Endoglucanase; Endo 3A; formation; localised; variation;
KW colour density; surface; dye; fabric; family 5; cellulose;
KW hydrolysis; p-nitrophenyl-beta-1,4-cellobioside; stone wash;
KW blue jeans; back staining.
XX
OS Bacillus lautus.
XX
PN WO9709410-A1.
PD 13-MAR-1997.
XX
PF 03-SEP-1996; 96WO-DK00364.
XX
PR 08-SEP-1995; 95DK-0000993.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Fich M, Onishi M, Schulein M, Toft AH;
XX WPI; 1997-192888/17.
XX
PT Localised variation of colour density in the surface of a dyed
cellulosic fabric - uses cellulase compsn. able to hydrolyse

PT p-nitrophenyl -beta-1,4-cellobioside
 XX Disclosure: Pages 15-17; 23pp; English.
 XX
 CC The present sequence is the corrected version of the incorrect
 CC Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A described in
 CC WO9110732. Endo 3 can be used in novel method of forming localised
 CC colour density variation on the surface of a dyed cellulosic
 CC fabric. The method comprises agitating the fabric in an aqueous
 CC medium (pH 6.5 to 9.0) containing a family 5 cellulose,
 CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
 CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
 CC abrading agent or cellulose having abrading activity. Each
 CC cellulase displays 30 % or more of its maximum activity at pH 7.
 CC The process is useful to provide a stone washed look to blue jeans
 CC without back staining.
 XX
 XX Sequence 551 AA;
 SQ

Query Match 45.1%; Score 374; DB 18; Length 551;
 Best Local Similarity 47.1%; Pred. No. 7.2e-29;
 Matches 72; Conservative 33; Mismatches 42; Indels 6; Gaps 4;

Qy 2 SCGVKVOYKKNDSAPCDNQIKPGQLVNTGSSVDLSTVTVRYWFTDRGGSTLVYNCW 61
 Db 402 TGNLVQYKVGDTSATDNQMKPSNKNKNGTTPVNLGKLRYFTKD-GTADMSASFDW 460
 Qy 62 AAMCGNIRASFGSVNPATPTADTYLQLSFT--GGTLAAGSGTGEIONRVKNSDWSNFDE 119
 Db 461 AQIGASVNSAAF--ANFTGSWTDYIVVELSFAGSGSIPAGGTGDIQLRMKYKTDSNFNE 518
 Qy 120 TNDYSY-GTNTAFQDWTKVTVYVNGRLVMGTEP 151
 Db 519 ANDYSYDCAKTAYADWNRVTLHQNGTLVWGTP 551

RESULT 2
 AARI3227
 ID AARI3227 standard; Protein: 700 AA.
 XX
 AC AARI3227;
 XX
 DT 14-OCT-1991 (first entry)
 XX
 DE Novel endoglucanase.
 XX
 KW Cellulase activity: detergent.
 XX
 OS Bacillus spp. NCIMB 40250.
 XX

Key Location/Qualifiers
 FT Peptide 1..31
 FT /label= signal sequence
 FT Protein 32..700
 FT /label= mature endoglucanase
 FT Cleavage-site 31..32
 XX
 PN WO9110732-A.
 XX
 XX
 PD 25-JUL-1991.
 XX
 XX
 PF 18-JAN-1991; 91WO-DK00013.
 XX
 XX
 PR 19-JAN-1990; 90DK-0000164.
 XX
 XX (NOVO) NOVO NORDISK A/S.
 PA
 XX Jorgensen PL, Schulein M, Hansen C;
 PI
 XX WPI: 1991-238020/32.
 DR
 DR N-PSDB; AAQ13001.
 XX
 PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an

PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
 PT fabrics.
 XX
 PS Claim 1; Page 80; 96pp; English.
 XX
 CC The enzyme is produced by a strain of Bacillus spp. NCIMB 40250
 CC and exhibits an endoglucanase activity of at least 10 (pref. at
 CC least 25) carboxymethyl cellulose (CMC) endoase units per mg total
 CC protein under alkaline conditions. It is especially useful as a
 CC cellulolytic agent and has been found to be more stable during
 CC washing (60 mins. at 40 deg.) in the presence of conventional
 CC detergents than a commercial cellulase preparation. It may also
 CC show increased storage stability in liq. detergents contg.
 CC proteases. The sequence was deduced from the DNA (AAQ13001).
 CC See also AARI3228 and AARI3229.
 XX
 XX Sequence 700 AA;
 SQ

Query Match 45.1%; Score 373.5; DB 12; Length 700;
 Best Local Similarity 49.7%; Pred. No. 1.1e-28;
 Matches 77; Conservative 32; Mismatches 39; Indels 7; Gaps 5;

Qy 1 VSGGVKVOYKKNDSAPCDNQIKPGQLVNTGSSVDLSTVTVRYWFTDRGGSTLVYNC 59
 Db 549 VNSDLVQYKDGDRNNATDNQIKPHFNQNKCTSPVDLSSLTRYVFTKD-SSAAMNGW1 607
 Qy 60 DWAAAGCGNIRASFGSVNPATPTADTYLQLSFT--GGTLAAGSGTGEIONRVKNSDWSNF 117
 Db 608 DWAKGGSGNIQISFGNHNGA--DSDTYAEFGSGAGSIAEGGQSGEIQLRMSKADWSNF 665
 Qy 118 DETNDYSY-GTNTAFQDWTKVTVYVNGRLVMGTEP 151
 Db 666 NEANDYSFDGAKTAYIDWDRVTLYODGQLVWGIEP 700

RESULT 3
 AARI5625
 ID AARI5625 standard; Protein: 782 AA.
 XX
 AC AARI5625;
 XX
 DT 17-MAR-1992 (first entry)
 XX
 DE Cellulase AE-1.
 XX
 KW Detergents; pharmaceuticals; deinking; carboxymethylcellulose.
 XX
 OS Aeromonas strain no. 212.
 XX

JP03251174-A.
 XX
 PD 08-NOV-1991.
 XX
 PF 28-FEB-1990; 90JP-0045465.
 XX
 PR 28-FEB-1990; 90JP-0045465.
 XX
 XX (OJIP) OJI PAPER KK.
 XX
 XX WPI: 1991-373412/51.
 DR
 DR N-PSDB; AAQ15178.
 XX

Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -
 of opt. pH when carboxymethylcellulose is used as substrate.
 PT
 XX
 XX Claim 2; Fig 3; 8pp; Japanese.
 XX

The sequence was deduced from the gene which was sequenced from
 CC plasmid, pAEC 1, prepd. by ligating chromosomal DNA contg. the
 CC gene (obtd. from Aeromonas) into pUC18. The protein has a mol. wt.
 CC of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-
 CC methylcellulose is the substrate. The N-terminal sequence: GIHADT-
 CC has been confirmed by Edman degradation. The gene can be used to

CC produce recombinant enzyme which is used for the effective utilis-
CC ation of biomass resources and the mfr. of pharmaceuticals and
CC foodstuffs, and also for the detergent and deinking of waste paper.
XX
SQ Sequence 782 AA:

Query Match 45.0%; Score 373; DB 12; Length 782;
Best Local Similarity 45.8%; Pred. No. 1.4e-28;
Matches 70; Conservative 30; Mismatches 49; Indels 4; Gaps 3;

QY 2 SCGVKVOYKNNDSAPCDNOIKPCLQLVNTGSSVDLSTVTVRVWFTRDGSSTLVNCDW 61
DB 631 SCDLAVOYKGTGDTNAADNOKPHFNIVNKGAAVPLSELRLRYFTAD-GNDOLQYNCW 689

QY 62 AAMCCGNIRASFGSVNPATPTADTYLQLSF--TGGTLAAGSGTGEIQNRVYKSDWSNFDE 119
DB 690 AMVCCSNLNCFAVKMNPGRKANADTYLEITFKAAGSLQAPACKRGISQTRNHAGNWANLNE 749

QY 120 TNDYSYG-TNTAFODWTKVTYVYVNGRLVMGTEP 151
DB 750 SNDYSFDPTKTAVANNKERVTLYHNGTLVFCNEP 782

RESULT 4
AAG63962
ID AAG63962 standard; Protein: 1352 AA.
AC AAG63962;
XX
DT 29-OCT-2001 (first entry)
XX
DE Amino acid sequence of xyloglucanase enzyme.
XX
KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW cellulosic fiber; textile scouring.
XX
OS Paenibacillus polymyxa.
XX
PN WO200162903-A1.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-DK00116.
XX
PR 24-FEB-2000; 2000DK-0000291.
XX
PA (NOVO) NOVOZYMES AS.
XX
PI Schnorr K, Jorgensen PL, Schuelein M;
XX
DR WPI; 2001-522819/57.
DR N-PSDB; AAH75059.
XX
PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT useful for detergent compositions, and textile or cellulose fiber
PT processing industries -
XX
PS Example 3; Page 81-85; 97pp; English.
XX
CC The present sequence represents a xyloglucanase of the invention. The
CC specification describes a xyloglucanase enzyme belonging to family 44
CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
CC detergent compositions and prevents binding of certain soils to the
CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
CC room temperature and has a half life of more than 50 days when incubated
CC in a full formulated liquid detergent at 30 plus degrees celcius. The
CC enzyme is used in detergent compositions, textile industry for improving
CC the properties of cellulosic fibers, yarn, and woven or non-woven
CC fabrics, preferably in textile scouring process, and in cellulose fiber
CC processing industry for rattling of fibers e.g. hemp, jute, flax, and
CC linen.
XX

SQ Sequence 1352 AA:

Query Match 44.5%; Score 369; DB 22; Length 1352;
Best Local Similarity 44.7%; Pred. No. 7.3e-28;
Matches 68; Conservative 33; Mismatches 47; Indels 4; Gaps 3;

QY 4 GVKVQYKNNDSAPCDNOIKPCLQLVNTGSSVDLSTVTVRVWFTRDGSSTLVYNCUWAA 63
DB 1200 GLLLQYRTADTKVNDHNLNFOFIVNKGTTSPINELKIRYYTIDGDREQ-TFNCDYAT 1258

QY 64 MCGCNIRASFGSVNPATPTADTYLQLSFT--GGTLAAGSGTGEIQNRVYKSDWSNFDETN 121
DB 1259 LSCSKLNGKLVKMEKATGADYVLEVSFNSDAGVLAPCGSGTDIOTRIHKTDMSYNESD 1318

QY 122 DYSY-CTNTAFODWTKVTYVYVNGRLVMGTEPS 152
DB 1319 DYSYKGTQTSFADHPKVTLYHNGVLVMGTEPT 1350

RESULT 5
AAU98063
ID AAU98063 standard; Protein: 677 AA.
AC AAU98063;
XX
DT 12-AUG-2002 (first entry)
XX
DE Bacillus subtilis spore coat protein cotG-CMCase fusion protein.
XX
KW Spore coat protein; spore surface; delivery system; bioconversion;
KW recombinant live vaccine; resistance property; protein microarray;
KW whole cell absorber; secretion signal; target signal; cell membrane;
KW surface display motif; cotG; CMCase; carboxymethyl cellulase.
XX
OS Chimeric - Bacillus subtilis.
OS Synthetic.
XX
PN WO200246388-A1.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001WO-KR02124.
XX
PR 08-DEC-2000; 2000KR-0074835.
XX
PA (GENO-) GENOFOCUS CO LTD.
XX
PI Pan J, Choi S, Jung H;
XX
DR WPI; 2002-463830/49.
DR N-PSDB; ABK53202.
XX
PT Displaying a protein of interest on spore surface, by transforming a
PT host cell with a vector comprising a gene encoding a spore coat protein
PT and a gene encoding a protein of interest, useful for preparing protein
PT microarray -
XX
PS Example 4; Page 150; 158pp; English.
XX
CC The present invention relates to a new method for displaying a protein
CC of interest on a spore surface. The method of the invention involves
CC preparing a vector comprising a gene construct containing a gene
CC encoding a spore coat protein and a gene encoding a protein of interest,
CC transforming a host cell with the vector, displaying the protein of
CC interest on a surface of a spore of the host cell, and recovering the
CC spore. The method is useful for producing a delivery system for
CC recombinant live vaccine. The method is also useful for improving a
CC protein of interest using a resistance property of spore, in
CC bioconversion reactions, for preparing protein microarray, for
CC producing an antibody to antigen in vertebrates, and for preparing a
CC whole cell absorber. The spore coat proteins used circumvent a necessity
CC for passage across cell membrane, so that they do not need secretion
CC signal and target signal which are prerequisites of surface display

CC and galactoglucmannans. The mannanase protein, or preparations
 CC containing it, are used to improve properties of cellulosic or
 CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods (feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.

XX Sequence 476 AA;

Query Match 43.1%; Score 357.5; DB 21; Length 476;
 Best Local Similarity 42.3%; Pred. No. 2.7e-27;
 Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVWFTRDGSSTLVYVNC 60
 DB 314 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTLYRYVTVDGOKDQTFW-CD 372
 QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIONRVNK 111
 DB 373 HAALIGNSGYNGITSNVKGTFVKMSSTNNADTYLEISFTGCTLEPCAHV-QIOGRFAK 431
 QY 112 SWSNFDFTNDYSYGTNTAFQDWTKVTVYVNGRLVWGTEPSGT 154
 DB 432 NDWSNYTQSDNYSFKRSQFVWDQVDTAYLNGVLVWGKEPGGS 474

RESULT 8
 AAY28850

ID AAY28850 standard; Protein: 493 AA.

XX AC AAY28850;

XX DT 17-JAN-2000 (first entry)

XX DE Pectate lyase-linker-CBD fusion protein.

XX KW Pectate lyase-linker-CBD fusion protein; ATCC 14580; plasmid pMB914;
 XX KW cellulose binding domain; CBD; linker; transformed; detergent;
 XX KW cellulosic fibre; yarn; degradation; recycled waste paper; animal feed;
 XX KW paper-making pulp; retting process; processing; wine; juice.

XX OS Chimeric - Bacillus licheniformis.
 XX OS Chimeric - Clostridium thermocellum.

XX PN WO9927083-A1.

XX PD 03-JUN-1999.

XX PF 24-NOV-1998; 98WO-DK00514.

XX PR 24-NOV-1997; 97DK-0001344.

XX PR 06-MAY-1998; 98US-0073684.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;

XX XX WPI: 1999-610578/52.

XX DR N-PSDB; AAX90978.

XX XX New isolated pectate lyase enzymes -

XX XX Example 4; Page 85-86; 93pp; English.

XX XX The present sequence is a Pectate lyase-linker-CBD fusion protein.

CC Plasmid pMB914 was constructed using pectate lyase gene from
 CC Bacillus licheniformis, ATCC 14580 and cellulose binding domain (CBD)
 CC gene from Clostridium thermocellum YS fused with a linker molecule.
 CC Bacillus subtilis was transformed with plasmid pMB914 for expression of
 CC the fusion protein. Pectate lyase can be used in detergent compositions,
 CC for cleaning hard surfaces, for machine treatment of fabrics, for
 CC improving the properties of cellulosic fibres, yarn, woven or non-woven
 CC fabric, for the degradation of plant material e.g. recycled waste paper,
 CC mechanical paper-making pulps or fibres subjected to retting process, for
 CC preparing animal feed and for processing wine or juice.

XX SQ Sequence 493 AA;

Query Match 43.1%; Score 357.5; DB 20; Length 493;
 Best Local Similarity 42.3%; Pred. No. 2.8e-27;
 Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVWFTRDGSSTLVYVNC 60
 DB 331 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTLYRYVTVDGOKDQTFW-CD 389
 QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIONRVNK 111
 DB 390 HAALIGNSGYNGITSNVKGTFVKMSSTNNADTYLEISFTGCTLEPCAHV-QIOGRFAK 448
 QY 112 SWSNFDFTNDYSYGTNTAFQDWTKVTVYVNGRLVWGTEPSGT 154
 DB 449 NDWSNYTQSDNYSFKRSQFVWDQVDTAYLNGVLVWGKEPGGS 491

RESULT 9
 AAY43218

ID AAY43218 standard; Protein: 493 AA.

XX AC AAY43218;

XX DT 13-JAN-2000 (first entry)

XX DE Pectate lyase CBD fusion protein sequence.

XX KW Pectate lyase; polysaccharide lyase; enzyme; pectin degradation;
 XX KW polygalacturonide; detergent composition; hard surface treatment;
 XX KW cellulosic fibre; plant material degradation; recycled waste paper;
 XX KW mechanical paper-making pulp; wine processing; cellulose binding domain;
 XX KW ClpB.

XX OS Clostridium thermocellum.

XX OS Bacillus sp.

XX OS Synthetic.

XX PN WO9927084-A1.

XX PD 03-JUN-1999.

XX PF 24-NOV-1998; 98WO-DK00515.

XX PR 24-NOV-1997; 97DK-0001343.

XX PR 24-NOV-1997; 97DK-0001344.

XX PR 06-MAY-1998; 98US-0073684.

XX PR 02-NOV-1998; 98US-0184217.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Andersen LN, Schuelein M, Lange NEK, Bjornvad ME, Moller S;

XX PI Glad SOS, Kauppinen MS, Schnorr K, Kongsbaek L;

XX DR WPI: 1999-610579/52.

XX DR N-PSDB; AAZ31562.

XX XX New isolated pectate lyase enzymes -

XX XX Claim 26; Page 106-108; 113pp; English.

PT assays, reactors, delivery etc.
 XX Claim 12: Page 34-35; 53pp: English.
 PS A modified cellulose binding domain (CBD) or fraction of it, linked
 CC to a hapten moiety via one or more Cys or Lys residues and complexes
 CC comprising the biotinylated CBD and a biotin-binding molecule
 CC selected from modified or unmodified avidin or streptavidin or an
 CC anti-biotin antibody, can be used in affinity chromatography, cell
 CC separation, cell, protein and enzyme immobilisation, selective
 CC removal of biological materials, multi-enzyme reactors, signal
 CC immunoassays and drug (e.g. pesticide) delivery. The CBD molecule
 CC can be modified with the hapten without affecting its high affinity
 CC for cellulose, or chitin.
 XX
 XX Sequence 167 AA:
 Query Match 43.0%; Score 356.5; DB 17; Length 167;
 Best Local Similarity 42.3%; Pred. No. 8.6e-28;
 Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;
 QY 1 VSGGVKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYVNC 60
 DB 5 VSGNLKVEFYNSNPSTTNSINPQFKVNTGSSAIDLSKLTLYRYTVVVDGQKDTFW-CD 63
 QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIQNRVK 111
 DB 64 HAAIIGSNGSYNGITSNVKCTFVKMSSSTNNADTYLEISFTGCTLEPGAHV-QIOGRFAK 122
 QY 112 SDWSNFDENDSYGTNTAFQDWTKVTVYVNGRLVNGTEPSGT 154
 DB 123 NDWSNVTOSNDYSFKXSOFVEMDQVATYLVNGVLVNGKEPGGS 165
 RESULT 12
 AAW15238
 ID AAW15238 standard; protein; 531 AA.
 XX
 AC AAW15238:
 XX
 DT 28-JAN-1998 (first entry)
 DE Scaffoldin protein from Clostridium thermocellum.
 XX
 KW Dockerin; CelB; CipA; scaffoldin; cellulose binding domain;
 KW chromatographic separation; soluble substrate modification; CBD;
 KW multi-enzyme delivery system; animal feed; paper production;
 KW plant protection; pest control.
 XX
 OS Clostridium thermocellum.
 XX
 FH Key Location/Qualifiers
 FT Region 1..153
 FT /label= internal_repeat_element_1
 FT Region 154..306
 FT /label= internal_repeat_element_2
 FT Domain 239..531
 FT /label= cellulose_binding
 FT /note= "only 60% of the CBD is present"
 XX
 DN W09714789-A2.
 XX
 XX 24-APR-1997.
 PD
 XX 16-OCT-1996; 96WO-US16485.
 PF
 XX 17-NOV-1995; 95US-0559968.
 PR
 PR 17-OCT-1995; 95US-0005701.
 XX
 PA (GEMV) GENENCOR INT INC.
 XX
 PI Bott RR, Clarkson KA, Fowler T, Liu C, Ward M, Xia H;
 XX

DR WPI: 1997-245106/22.
 XX Composition with enzymes non-covalently bound to a peptide backbone
 PT - used as a multi-enzyme delivery system, e.g. in food processing,
 PT textiles and pest control
 XX Claim 10: Fig 6; 29pp: English.
 PS
 XX This protein sequence represents a scaffoldin based on the CipA protein
 CC of Clostridium thermocellum. The scaffoldin protein is used in a novel
 CC composition that comprises at least 2 enzymes non-covalently bound to a
 CC peptide backbone (i.e. present sequence). The scaffoldin comprises a
 CC number of internal repeating units and at least one cellulose binding
 CC domain (CBD). The CBD may be altered to modify its affinity for
 CC cellulose, which may be desirable where cellulose binding would be
 CC disadvantageous. The enzyme is bound to the scaffoldin by a dockerin
 CC region of the enzyme, which binds to a repeating element of the
 CC scaffoldin. The dockerin is preferably Cels (AAW15237) or CelD
 CC (AAW15236). The composition can be used in reducing allergenicity,
 CC producing synergistic effects, and facilitating selective modification
 CC of substrate. By taking advantage of the cellulose binding domain of the
 CC complex, the complex could be immobilised for use in chromatographic
 CC separations or for soluble substrate modification. By adding the
 CC scaffoldin domain, it is possible to recover enzymes, or to quantify
 CC the amount of an enzyme in a solution. The composition could also be
 CC used in a multi-enzyme delivery system which could be used in the food
 CC industry, in food processing, animal feed, textiles, bioconversion,
 CC pulp and paper production, plant protection and pest control, as a wood
 CC preservative, topical lotions, and biomass conversions.
 XX
 XX Sequence 531 AA:
 Query Match 42.9%; Score 355.5; DB 18; Length 531;
 Best Local Similarity 42.3%; Pred. No. 4.9e-27;
 Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;
 QY 1 VSGGVKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYVNC 60
 DB 312 VSGNLKVEFYNSNPSTTNSINPQFKVNTGSSAIDLSKLTLYRYTVVVDGQKDTFW-CD 370
 QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIQNRVK 111
 DB 371 HAAIIGSNGSYNGITSNVKCTFVKMSSSTNNADTYLEISFTGCTLEPGAHV-QIOGRFAK 429
 QY 112 SDWSNFDENDSYGTNTAFQDWTKVTVYVNGRLVNGTEPSGT 154
 DB 430 NDWSNVTOSNDYSFKXSOFVEMDQVATYLVNGVLVNGKEPGGS 472
 RESULT 13
 AAW43108
 ID AAW43108 standard; Protein; 1853 AA.
 XX
 AC AAW43108:
 XX
 DT 16-OCT-1998 (first entry)
 DE C. thermocellum cellulosome integrating protein.
 XX
 KW Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
 KW cohesin domain; catalytic subunit; scaffold subunit; SdBA; synergistic;
 KW cellulosome integrating protein; scaffoldin dockerin binding protein.
 XX
 OS Clostridium thermocellum.
 XX
 PN FR2748479-Al.
 XX
 PD 14-NOV-1997.
 XX
 XX 10-MAY-1996; 96FR-0005854.
 PF
 XX 10-MAY-1996; 96FR-0005854.
 PR
 XX

PA (INSP) INST PASTEUR.
XX Beguin P, Leibovitz E;
PI WPI: 1998-011569/02.
XX DR N-PSDB: AAT86625.
XX
PT Cellulase proteins with cohesin or dockerin type II domains - useful
PT for potentiating the activity of multiprotein enzyme complexes
XX
PS Disclosure; Page 47; 60pp; French.
XX
CC Multimeric protein, especially enzymatic, complexes are held together
CC by protein-protein interactions between domains designated dockerins
CC and cohesins, which are found on the catalytic and scaffold subunits
CC respectively. An example of such a complex is the cellulose degrading
CC protein complex from Clostridium thermocellum, known as the cellulosome.
CC This complex comprises around 15 proteins including endoglucanases,
CC cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which
CC interact with a central "scaffold" protein designated the cellulosome
CC integrating protein (Cip; see AAW43108). The catalytic subunits
CC interact with the Cip subunit via conserved 23 amino acid dockerin
CC domains. Cip has been shown to contain 9 copies of a cohesin domain.
CC The invention relates to the isolation of proteins binding to a novel
CC dockerin type domain found in the C-terminal portion of Cip. The new
CC domain is designated a type II dockerin domain (as compared to the type
CC I domain found on the catalytic subunits of the cellulosome). The type
CC II dockerin domain has some sequence similarity to the type I dockerins
CC but is unable to bind type I cohesin domains.
CC The sequence presented here is the cellulosome integrating protein
CC (CIP-A) to which the proteins SdhA, OlpB and ORF2p (AAW43105-W43107) can
CC bind at the C-terminal dockerin type II domain (location not given in
CC specification). The novel type II dockerin and cohesin domains can be
CC used in complexes, especially enzyme complexes, to potentiate their
CC catalytic actions in a synergistic manner.
XX
SQ Sequence 1853 AA;

Query Match 42.9%; Score 355.5; DB 19; Length 1853;
Best Local Similarity 42.3%; Pred. No. 2.5e-26;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKNDNSAPGCDNQIKPGQLVNTGSSSVLDLSTVTVRVWFTRDGSSSTLVYVNC 60
DB 1 VSGGVKVOYKNDNSAPGCDNQIKPGQLVNTGSSSVLDLSTVTVRVWFTRDGSSSTLVYVNC 60
QY 61 WAAM-----CGGNIRASFGSVNPATPTADTYLQLSFTGTT--LAAGSGTGEIQNRVVK 111
DB 424 HAAIIGSNGSYNGITSNVKGTFFVKMSSTNNADTYLEISFTGTTLEPCAIV-QIQGRFAK 482
QY 112 SDWSNFDENYSGTNTATFODWTKVTVVYVNGRLVWGTEPSGT 154
DB 483 NDWSNYTQSDNYSFKSASQFVENDQVATAYLNGVLVWGKEPGGS 525

RESULT 14
AAV13494
ID AAV13494 standard; Protein: 616 AA.
XX
AC AAV13494;
XX
DT 30-JUL-1999 (first entry)
XX
DE Truncated cellulase Cel E3/B5.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
OS Unidentified.
XX
PN EP921188-A2.
XX

PD 09-JUN-1999.
XX
PF 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
PA (CLRN) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI: 1999-315403/27.
DR N-PSDB: AAX55660.
XX
PT New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
XX
PS Claim 7; Page 42-43; 65pp; English.
XX
CC The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492. Cel E1 extends
CC from amino acid Y39 to D481. Cel E1/2 extends from Y39 to G635. Cel
CC E1/2/3 extends from Y39 to G812. Cel E6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAY13493. Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or
CC remove staining, backstaining or graying, for use on cellulosic
CC materials including cotton-containing fabrics. They are especially useful
CC for preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using
XX non-truncated cellulase compositions.
XX
SQ Sequence 616 AA;

Query Match 42.1%; Score 349; DB 20; Length 616;
Best Local Similarity 43.8%; Pred. No. 2.6e-26;
Matches 71; Conservative 26; Mismatches 49; Indels 16; Gaps 4;

QY 1 VSGGVKVOYKNDNSAPGCDNQIKPGQLVNTGSSSVLDLSTVTVRVWFTRDGSSSTLVYVNC 60
DB 1 MGSQKVLKYNKNETSASTGSRPWFKIVNGSSSVLDLSTVRKIRVWFTVDDGKPOSAV-CD 59
QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQLSFTGTT--LAAGSGTGEIQNRVVKSDWSNFD 118
DB 60 WAQIGASNVTFNFVKLSGVSQADYILEVGFSSGAGLOLPGCKDTGDIQVRFKNKNDWSYN 119
QY 119 ETNDYS-----YGTNTAFODWTKVTVVYVNGRLVWGTEPSG 153
DB 120 QADDMSWLQSMNTNYGEN-----AKVTLYVDGLVWGTEPGG 155

RESULT 15
AAE16325
ID AAE16325 standard; Protein: 616 AA.
XX
AC AAE16325;
XX
DT 26-MAR-2002 (first entry)
XX
DE Active cellulase hybrid protein, E3/B5.
XX
KW Active cellulase protein; alkaliphilic; textile processing; proteinase;
KW detergent additive; stonewashed appearance; cotton-containing denim;
KW CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein.
XX
OS Unidentified.
XX

A: Note: sequence extracted from NCBI backbone (NCBIP:104605)
C: Function:
A: Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A: Pathway: cellulose degradation
C: Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 45.1% Score 373.5 Dh 2 Length 700;
Best Local Similarity 49.7% Pred. No. 7.8e-26;
Matches 77: Conservative 32: Mismatches 39: Indels 7: Gaps 5:

QY 1 VSGGVKVVQNND-SAPGDNQIKPGLVNTGTSSVDLSLTVTYRYWFTTRDGGSTLVINC 59
 : : | | | | : : | | | | : : | | | | : : | | | | : :
Db 549 VNSDLVVQYKGDRNNATDNQIKPHENIQNKGTSPVDLSLLTRYVFTKD-SSAAANGWI 607

QY 60 DWAAGCCNTRASFGSVNPATPTADTYQLSLFT--GCTLAAGCGTGEIQRNVKSDWSNF 117
 : : | | | | : : | | | | : : | | | | : : | | | | : :
Db 608 DWALGGSNIOTISFNGINGA--DSDTYAELGFSSGACSI AEGSGSIEIQRMSKADWSNF 665

QY 118 DETNDYSV-CNTAFQDMTKVTIVYNGRLVMGTEP 151
 : | | | | : | | : | | : | | | | | | | | | |
Db 666 NEANDYSFDCAKTAYIDMRVTLYQGOLVWGIEP 700

RESULT 3
G69593
cellulase (EC 3.2.1.4) bgIC precursor - Bacillus subtilis
N: Alternate names: endo-1,4-beta-galactanase
C: Species: Bacillus subtilis
C: Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C: Accession: G69593; A26114; I40353; S24239; S49103; I39803
R: Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allison, G.; Azevedo, V.; Berter
C.: Aron, S.; Broutelle, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A: Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols,
A: Authors: Lauber, R.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel
y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetlelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A: Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Setor
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpfstra, P.; Toqnoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A: Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A: Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A: Reference number: A69580; MID: 98044033; PMID: 9384377
A: Accession: G69593
A: Status: nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-508 <KUN>
A: Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13696.1; PID:e1183471;
A: Experimental source: strain 168
R: Mackay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, F.; Seili
Nucleic Acids Res. 14, 9159-9170, 1986
A: Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
A: Reference number: A26114; MID: 87066783; PMID: 3024130
A: Accession: A26114
A: Molecule type: DNA
A: Residues: 10-508 <MAC>
A: Cross-references: GB:X04689; NID:g39823; PIDN:CAA283392.1; PID:g39824
A: Experimental source: strain PAP115
A: Note: part of this sequence, including the amino end of the mature form, was confirmed
R: Lindahl, V.; As, K.; Tronsmo, A.
Antonie Van Leeuwenhoek 66, 327-332, 1994
A: Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis Q
A: Reference number: I40353; MID: 95225656; PMID: 7710280
A: Accession: I40353
A: Status: preliminary; translated from CB/EMBL/DDBJ
A: Molecule type: DNA
A: Residues: 10-291, N' 293-508 <LIN1>
A: Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
R: Lindahl, V.; As, K.
submitted to the EMBL Data Library, June 1992

[illegible]

R:Gerngross, U.T.; Demain, A.L.
submitted to the EMBL Data Library, January 1993
A:Reference number: S36859
A:Accession: S36859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1854 <GER>
A:Cross-references: EMBL:L08665
R:Gerngross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.
Mol. Microbiol. 8, 325-334, 1993
A:Title: Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal
A:Reference number: S33527; MUID:93302508; PMID:8316083
A:Accession: S33527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1615,1617-1854 <GE2>
A:Cross-references: EMBL:L08665
R:Fujino, T.; Beguin, P.; Aubert, J.P.
FEMS Microbiol. Lett. 94, 165-170, 1992
A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that b
A:Reference number: S25767
A:Accession: S25767
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615,1617-1854 <FUJ>
A:Cross-references: EMBL:X67406
R:Bequin, P.
submitted to the EMBL Data Library, August 1992
A:Reference number: S28659
A:Accession: S28659
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615,1617-1768, 'R', 1770-1854
A:Cross-references: EMBL:X67406
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: 218847; MUID:93209931; PMID:8458832
A:Accession: T18261
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1821-1854 <FU2>
A:Cross-references: EMBL:X67506; NID:9296879; PID:9296880; PIDN:CAA47840.1
C:Genetics:
A:Gene: cipA
Query Match 42.9%; Score 355.5; DB 2; Length 1854;
Best Local Similarity 42.3%; Pred. No. 1e-23;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;
Qy 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYWFTRDGGSTLVYKNC 60
Db 365 VSGNLKVEFYNSNPSTTNSINPQFKVTNTGSSAIDLKSLTRYTYTVDGQKQDTFW-CD 423
Qy 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGTTLAAGSGTGEIQNRVKN 111
Db 424 HAAIIGSNGSYNGITSNVKGTEVRKMSSTNNADTYLEISFTGTTLEPGAHV-OIQGRFAK 482
Qy 112 SDWSNFDENDSYGTNTAFQDWTKVTYVYVNGRLVWGTEPSGT 154
Db 483 NDWSNYTOSNDYSFKSASQFVEMDQVTAYLNGVLVWGKEPGGS 525
RESULT 9
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139; PMID:1476429

A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GIB>
A:Cross-references: GB:L01257; NID:9144290; PIDN:AAA71887.1; PID:g144291
R:Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPRQHQHRQ' <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
A:Note: the authors translated the codon CAC for residue 262 as Glu
A:Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
Query Match 42.5%; Score 352; DB 2; Length 1331;
Best Local Similarity 45.8%; Pred. No. 1.4e-23;
Matches 71; Conservative 30; Mismatches 50; Indels 4; Gaps 3;
Qy 3 GGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYWFTRDGGSTLVYKNCWA 62
Db 568 GOIKVLYANKETNSTTNTIRPWLKVYVNSGSSSIDLSRVTIRYWTVDGERAQSAIS-DWA 626
Qy 63 AMGCCNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFDCT 120
Db 627 QIGASNVTFKFKVLSVSSVSGADYILEIGKSGAGLOPKDTGCTOIRFNKSDWSNNOG 686
Qy 121 NDYSY-GTNTAFQDWTKVTYVYVNGRLVWGTEPSGT 154
Db 687 NDWSIQSNTSYGENEKVTAYIDGVLVWGQEPSGT 721
RESULT 10
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: Cella, another gene coding for a multidomain cellulase from the extreme therm
A:Reference number: 218698; MUID:95336703; PMID:7612247
A:Accession: T17120
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A:Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1
R:Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C:Genetics:
A:Gene: cella
C:Keywords: glycosidase; hydrolase
Query Match 42.5%; Score 352; DB 2; Length 1742;
Best Local Similarity 45.5%; Pred. No. 2e-23;
Matches 71; Conservative 30; Mismatches 51; Indels 4; Gaps 3;
Qy 2 SGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYWFTRDGGSTLVYKNCW 61
Db 704 SQIKVLYANKETNSTTNTIRPWLKVYVNSGSSSIDLSRVTIRYWTVDGERAQSAIS-DW 762
Qy 62 AAMGCCNTRASFSGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFDCT 119

Db 763 AQICASNVTFFVKLSSSVSGADYYLEIGFKSGAGOLQPKDGTGEIQIRFNKDDWSNYNQ 822
 Qy 120 TNDYSY-CTNTAFQDWTKTVVYVNGRLVNGTPEPST 154
 Db 823 GNDWSIQSMTSYGENEKVTAYIDGVVLVWGQEPST 858

RESULT 11
 S02711
 cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
 N:Alternate names: endo-1,4-beta-glucanase
 N:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3
 C:Species: Caldocellum saccharolyticum
 C:date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
 C:Accession: S02711
 R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
 Nucleic Acids Res. 17, 439, 1989
 A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum enc
 A:Reference number: S02711; MUID:89098398; PMID:2789517
 A:Accession: S02711
 A:Molecule type: DNA
 A:Residues: 1-1039 <SAU>
 A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
 C:Genetics:
 A:Gene: celB
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-gluc
 A:Pathway: cellulose degradation
 C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-1039/Product: cellulase #status predicted <MAT>
 F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <X>

Query Match 41.9%; Score 347; DB 2: Length 1039;
 Best Local Similarity 45.5%; Pred. No. 3e-23;
 Matches 70; Conservative 30; Mismatches 50; Indels 4; Gaps

Qy 3 GGVKVQYKNNDSPGQNIKPGQLQVLVNTGSSSVLDLSTVTYVWFTRDGSGSTLVYNCDA 62
 Db 420 GQIKVLVYANKETNSTNTIRPLVKVNVSSSIDLSRVTIRYWTVDGERAQSAVS-DWA 478

Qy 63 AMCGGNIRAFSGVNPATPTATYTLQLSFTGT--LAAGSTGETIQNRVKNKSDWSNDET 120
 Db 479 QIGASNVTFFVKLSSSVSGADYYLEIGFKSGAGOLQPKDGTGEIQIRFNKDDWSNYNQ 538

Qy 121 NDYSY-CTNTAFQDWTKTVVYVNGRLVNGTPEPSG 153
 Db 539 NDWSIQSMTSYGENEKVTAYIDGVVLVWGQEPST 572

RESULT 12
 T31085
 xylanase - Caldicellulosiruptor sp.
 C:Species: Caldicellulosiruptor sp.
 C:date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
 C:Accession: T31085
 R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
 submitted to the EMBL Data Library, December 1997
 A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp
 A:Reference number: 220972
 A:Accession: T31085
 A:Molecule type: DNA
 A:Residues: 1-1779 <MOR>
 A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95322
 C:Genetics:
 A:Note: xynC

Query Watch 41.5%; Score 344; DB 2: Length 1779;
 Best Local Similarity 44.0%; Pred. No. 1.1e-22;
 Matches 70; Conservative 25; Mismatches 48; Indels 16; Gaps

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:20/320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 41.1%; Score 341; DB 2; Length 915;
Best Local Similarity 45.1%; Pred. No. 9e-23;
Matches 69; Conservative 30; Mismatches 50; Indels 4; Gaps 3;

Qy 3 CGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSGSTLVYNC DWA 62
Db 349 GOIKVLYANKETNTTIRPWLKVNSSGSSIDLRSVTIRYWTVDGERAQSAYS-DWA 407
Qy 63 AMCCGNTIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGETQNRVNKSDWSNFDET 120
Db 408 QIGASNVTFKFKLSSSVSGADYYLEICFKSGAGQLOQPKDTGELQIRFNKSDWSNYNQG 467
Qy 121 NDYSY-GTNTAFQDWTKVTYVNGRLVMGTEPS 152
Db 468 NDWSWLOSMTSYGENEKVTAYIDGVLVNGQEPS 500

RESULT 15
S39962
endoglucanase - Erwinia carotovora
C:Species: Erwinia carotovora
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S39962
R:Cooper, V.J.C.; Salmund, G.P.C.
Mol. Gen. Genet. 241, 341-350, 1993
A:Title: Molecular analysis of the major cellulase (CelV) of Erwinia carotovora: evidenc
A:Reference number: S39962; MUID:94067016; PMID:8246888
A:Accession: S39962
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-505 <COO>
A:Cross-references: EMBL:X76000; NID:9434941; PIDN:CAA53592.1; PID:9434942

Query Match 40.4%; Score 335; DB 2; Length 505;
Best Local Similarity 43.9%; Pred. No. 1.6e-22;
Matches 68; Conservative 32; Mismatches 47; Indels 8; Gaps 4;

Qy 2 SGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDG--GSSTLVYNC 59
Db 354 TGDVVLQYRNVNDNPSDDAIRMAVNIKTGSTPIKLSDLQVRYFHDGKPGCANLFV--- 410
Qy 60 DWAAMGCGNIRASFGSVNPATPTADTYLQLSFT--GCTLAAGSGTGETQNRVNKSDWSNF 117
Db 411 DWANVGPNNIIVTSTGTAASTDKANRYVLVTFSSGAGSLQPGAETGEVQVRIHAGDWSNV 470
Qy 118 DETNDYSYGTN-TAFQDWTKVTYVNGRLVMGTEP 151
Db 471 NETNDYSYGANVTSYANWKDITVHDKGTLVNGVEP 505

Search completed: November 13, 2002, 11:57:35
Job time : 9.19109 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:47:41 : Search time 3.54023 Seconds
(without alignments)
1804.220 Million cell updates/sec

Title: US-09-917-378-4

Perfect score: 829

Sequence: 1 VSGGVKQYKNDSDAPGDNQ.....TKTVVYVNGRLVWGTEPSC 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394	47.5	145	1 YCEA_PAE1A	P29718 paenibacill
2	373.5	45.1	700	1 GUN2_BACSU	P29719 paenibacill
3	370	44.6	499	1 GUN2_BACSU	P10475 bacillus su
4	368	44.4	499	1 GUN3_BACSU	P23549 bacillus su
5	365	44.0	499	1 GUN1_BACSU	P07983 bacillus su
6	357.5	43.1	772	1 CIPB_CLOTM	Q01866 clostridium
7	355.5	42.9	1853	1 CIPB_CLOTM	Q06851 clostridium
8	352	42.5	1331	1 MAND_CALSA	P22533 caldocellum
9	352	42.5	1742	1 GUN2_CALSA	P22534 caldocellum
10	347	41.9	1039	1 GUN2_CALSA	P10474 c endogluc
11	343.5	41.4	504	1 GUNV_ERWCA	Q59395 erwinia car
12	335	40.4	505	1 GUNV_ERWCA	Q47096 erwinia car
13	295.5	35.6	914	1 GUN2_CLOSR	P50900 clostridium
14	285	34.4	986	1 GUN2_CLOSR	P23659 clostridium
15	257	31.0	879	1 GUN1_CLOTM	Q02934 clostridium
16	225	27.1	1848	1 CBPA_CLOCL	P38058 clostridium
17	149.5	18.0	444	1 GUNV_ERWCA	Q59394 erwinia car
18	136.5	16.5	1045	1 GUNB_CELFI	P26225 cellulomona
19	101.5	12.2	636	1 GUN4_BACSU	P28622 bacillus sp
20	99	11.9	880	1 GUN4_THFVU	P26221 thermomonas
21	97.5	11.8	269	1 EXOK_RHIME	P33693 rhizobium m
22	89.5	10.8	725	1 GUN2_CLOCE	P37700 clostridium
23	81.5	9.8	582	1 HEMA_MUMPM	P11235 mumps virus
24	79.5	9.6	698	1 YMCA_ECOLI	P75882 escherichia
25	79.5	9.6	1200	1 ICEN_PSESY	P06820 pseudomonas
26	79	9.5	238	1 GUN2_PAEPO	P45797 paenibacill
27	78.5	9.5	1210	1 ICEN_PSEFL	P09815 pseudomonas
28	77.5	9.3	498	1 VNUC_TAUSS	P18073 influenza a
29	77.5	9.3	582	1 HEMA_MUMPM	P10866 mumps virus
30	77.5	9.3	1567	1 ICEN_XANCT	P18127 xanthomonas
31	77	9.3	237	1 GUNB_PAEPA	P23904 paenibacill
32	77	9.3	282	1 POR2_XENLA	P81004 xenopus lae
33	77	9.3	1006	1 BGAL_ASPNG	P29853 aspergillus

34	76	9.2	247	1 FLAI_THEVO	P57719 thermoplasm
35	76	9.2	1157	1 C9CA_BACTO	O45733 bacillus th
36	75.5	9.1	644	1 XYND_CELFI	P54865 cellulomona
37	74.5	9.0	488	1 PHB_ALCPA	P12625 alicigenes
38	74.5	9.0	498	1 VNUC_IABRA	P18069 influenza a
39	74.5	9.0	498	1 VNUC_IAPOW	P18071 influenza a
40	74.5	9.0	498	1 VNUC_IAME3	O07545 influenza a
41	74.5	9.0	582	1 HEMA_MUMPM	P19762 mumps virus
42	74	8.9	1148	1 ICEK_PSESX	O30611 pseudomonas
43	74	8.9	1244	1 SLAI_YEAST	P32790 saccharomyc
44	74	8.9	3462	1 RELN_RAT	P58751 rattus norv
45	73.5	8.9	242	1 GUB_BACSU	P04957 bacillus su

ALIGNMENTS

RESULT 1

YCEA_PAE1A

ID YCEA_PAE1A STANDARD; PRT; 145 AA.

AC P29718; 92276330; PubMed=1592807;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein in CELA 5' region (Fragment).

OS Paenibacillus lautus (Bacillus lautus).

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.

OX NCBI_TaxID=1401;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PL236;

RX MEDLINE=92276330; PubMed=1592807;

RA Hansen C.K., Joergensen P.L., Diderichsen B.;

RT "Cella from Bacillus lautus PL236 encodes a novel cellulose-binding

endo-beta-1,4-glucanase.";

RL J. Bacteriol. 174:3522-3531(1992).

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

[illegible]

RESULT 4	
GUN3_BACSU	STANDARD; PRT; 499 AA.
ID	GUN3_BACSU
AC	P23549;
DT	01-NOV-1991 (Rel. 20, 'Created)
DT	01-NOV-1991 (Rel. 20, Last sequence update)
DT	01-FEB-1995 (Rel. 31, Last annotation update)
DE	Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE	(Carboxymethyl-cellulase) (CMCase) (Cellulase).
BGLC.	
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=1423;	
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BSE616;
RC	MEDLINE=91299280; PubMed=1368694;
RA	Park S.H., Kim H.K., Pack M.Y.;
RT	"Characterization and structure of the cellulase gene of Bacillus
RT	subtilis BSP616.";
RL	Agric. Biol. Chem. 55:441-448(1991).
CC	- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC	linkages in cellulose.
CC	- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC	HYDROLASES).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC	or send an email to license@lsb-sib.ch).
CC	-----
EMBL:	D01057; BAA00859.1; ..
PIR:	JN0111; JN0111.
HSSP:	O85465; IA3H.
DR	InterPro: IPR001956; CBD_3.
DR	InterPro: IPR001547; GH_5.
DR	Pfam: PF00150; cellulase; 1.
DR	Pfam: PF00942; CBM_3; 1.
DR	ProDom: PD001947; CBD_3; 1.
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW	SIGNAL
FT	1 29
FT	CHAIN 30 499
FT	ACT_SITE 169 169
FT	ACT_SITE 257 257
FT	DOMAIN 350 499
FT	SEQUENCE 499 AA; 55169 MW; 2E821E3D8BBACA04 CRC64;
SQ	
	ENDOGLUCANASE.
	PROTON DONOR (BY SIMILARITY).
	NUCLEOPHILE (BY SIMILARITY).
	CELLULOSE-BINDING (BY SIMILARITY).

```

Query Match          44.4%; Score 368; DB 1; Length 499;
Best Local Similarity 43.6%; Pred. No. 6.7e-27;
Matches 65; Conservative 31; Mismatches 51; Indels 2; Gaps 1;

OY      4   GVKVOYKKNDAPGDNOIKPGLQLVNTGSSVDLSTVTYRVWFTRDGSSTLVYNCDHAA 63
         I : ||| : | : ||| : ||| : | : ||| : ||| : | : ||| :
DB     353 GISVOYRAGDGSMNSNQIRPOLQTKNNGNMTTVDLKDVTRYWYNAKNKGQNV--DCDYAQ 410
         : ||| : | : ||| : ||| : | : ||| : ||| : | : ||| :

OY     64   MCGGNIRASFSGSNPATDTADTYLQLSPGTCTLAAGSGTGQNRVNKSDMSNFDETDY 123
         : ||| : | : ||| : ||| : | : ||| : ||| : | : ||| :
DB    411 LCCGNVTVKFVTLHKPGOGADTYLELGFKNCTLPAGASTGNIQLRLHNDDNSNTAQSUDY 470
         : ||| : | : ||| : ||| : | : ||| : ||| : | : ||| :

OY    124   SYGTNTAFQDWTKVTVVYNGRLVMGCTEPS 152
         I : | : | : | : | : | : | : | : | : | : | : | : | :
DB    471 SFFKSNTETKTKKITLYDOGKLIMGTEPN 499

RESULT 5
GUN1_BACSU
ID      GUNI_BACSU        STANDARD:      PRT:      499 AA.
AC      P07983;
DT      01-AUG-1988 (Rel. 08, Created)
DT      01-NOV-1991 (Rel. 20, Last sequence update)
DT      01-OCT-1994 (Rel. 30, Last annotation update)
DE      Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
(Gallulase).
GN      BGIC OR GLD.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxId=1423;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=DLC;
RX      MEDLINE=87194581; Pubmed=3106328;
RA      Robson L.M., Chambliss G.H.;
RT      "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
RL      J. Bacteriol. 169:2017-2025(1987).
CC      -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
links in cellulose.
CC      -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES)
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC      EMBL; M16185; AAA22496.1; ALT_INIT.
DR      PTR; A26874; A26874.
DR      HSP; O85465; IA3H.
DR      InterPro; IPR001956; CBD_3.
DR      InterPro; IPR001547; GH_5.
DR      Pfam; PF00150; cellulase; 1.
DR      Pfam; PF00942; CBM_3; 1.
DR      ProDom; PD001947; CBD_3; 1.
DR      PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
FT      SIGNAL
FT      CHAIN
FT      CD
FT      ACT_SITE 169 499 ENDOGLUCANASE.
FT      ACT_SITE 257 257 PROTON DONOR (BY SIMILARITY).
FT      DOMAIN 350 499 NUCLEOPHILE (BY SIMILARITY).
FT      DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ      SEQUENCE 499 AA; 55187 MW; 339004EE95A63EE1 CRC64;

Query Match          44.0%; Score 365; DB 1; Length 499;
Best Local Similarity 44.7%; Pred. No. 1.3e-26;
Matches 67; Conservative 31; Mismatches 48; Indels 4; Gaps 2;

OY      4   GVKVOYKKNDAPGDNOIKPGLQLVNTGSSVDLSTVTYRVWFTRDGSSTLVYNCDHAA 62
         ||||| | : ||| : | : ||| : ||| : | : ||| : ||| : | : ||| :

```

```

Db 353 GVSVOYKAGDGRVNSQIRPOLHIKNNGNATVDLKDVYRWYVNVKNKGON----FDCDYA 409
QY 63 AMCCGNIASFSGVSNPATPTADTYLQLSFTGGTTLAAGSGTGEIQNRVNSKDSWNSFDETND 122
    |||||:|:::| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
Db 410 QMCCGNLTHKFTVYLLHKPKGADTYLELGFKTGTLSPGASTGNLQLKLNDDWSNYAQSOD 469
QY 123 YSYGNTAFQDWTKVTYVYVNGRLVWGTPEPS 152
    ||:|:|:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
Db 470 YSFFQSNTKTKTKITLYHQGLKIMCTEPN 499

RESULT 6
CIPA_CLOTH
ID CIPA_CLOTH STANDARD: PRT: 772 AA.
AC Q01866;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein sl/sl)
DE (Cellulose integrating protein B) (Fragment).
GN CIPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93146373; PubMed=1490597;
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,
  Gilbert H.J.;
RT "Identification of the cellulose-binding domain of the cellulosome
  subunit Sl from Clostridium thermocellum Ys."
RL FEMS Microbiol. Lett. 78:181-186(1992).
CC -!- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
  PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
  CELLULOYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
  REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
  PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.
CC -!- SUBCELLULAR LOCATION: CELL SURFACE.
CC -!- SIMILARITY: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement. (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch.)
CC
CC EMBL; X68233; CAA48312.1; -.
CC HSP; Q06851; INBC.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR002102; Cohesin.
CC InterPro; IPR002105; Dockerin_1.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00404; Dockerin_1; 2.
CC Pfam; PF00942; CBW_3; 1.
CC Pfam; PF00963; Cohesin; 3.
CC ProDom; PD001947; CBD_3; 1.
CC PROSITE; PS00018; EF_HAND; UNKNOWN_1.
CC PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN 1
FT REPEAT <1 607 3 X 147 AA APPROXIMATE REPEATS.
FT REPEAT <1 80 1 (PARTIAL).
FT REPEAT 94 240 2.
FT REPEAT 462 607 3.
FT DOMAIN 81 93 PRO/THR-RICH.
FT DOMAIN 241 272 PRO/THR-RICH.
FT DOMAIN 273 439 CELLULOSE-BINDING.
FT DOMAIN 440 461 PRO/THR-RICH.
FT DOMAIN 710 766 2 X 24 AA APPROXIMATE REPEATS.

```

```

FT REPEAT 710 733 2-1.
FT REPEAT 743 766 2-2.
SQ SEQUENCE 772 AA; 82491 MW; BBF06DE5E094FE10 CRC64;

Query Match 43.1%; Score 357.5; DB 1; Length 772;
Best Local Similarity 42.3%; Pred. No. 1.le-25;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY 1 VSGGVKVOYKNDSAPCDNOIKPCLQLVNTGSSVDLSLTVTYRWYFTRDGGSSFLVYNCD 60
    |||||:|:::| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
Db 277 VSNLKVFEYNSNPSTTNSINQFKVNTGSSAIDLSKLTLYRYTYVDGKQDTFW-CD 335
    |||||:|:::| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
QY 61 WAAM-----CGGNIRASFSGVSNPATPTADTYLQLSFTGGTTLAAGSGTGEIQNRVKN 111
    ||:|:|:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
Db 336 HAAIIGSNGSYNGCITSNVKCTFKVMSSTNNADTYLEISFTGCTLEPGAHV-OIOGRFAK 394
    |||||:|:::| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
QY 112 SDWSNFDENIDSYGNTAFQDWTKVTYVYVNGRLVWGTPEPSGT 154
    |||||:|:::| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
Db 395 NDWSNTQSDNYSFKRSQFVENDQVTAYLNGVLVWGTPEPGS 437

RESULT 7
CIPA_CLOTH
ID CIPA_CLOTH STANDARD: PRT: 1853 AA.
AC Q06851;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (Cellulosomal
  glycoprotein sl/sl) (Cellulose integrating protein A) (Cohesin).
GN CIPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 29-40 AND 42-43.
RX STRAIN-ATCC 27405 / DSM 1237;
RX MEDLINE=93302508; PubMed=8316083;
RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
  Demain A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cipa) encoding the
  cellulosomal SL-protein reveals an unusual degree of internal
  homology."
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RP SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE=93209931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
  cellulosomal scaffolding protein Cipa and a protein possibly involved
  in attachment of the cellulosome to the cell surface."
RL J. Bacteriol. 175:1891-1899(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE=97238934; PubMed=9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
  Frolow F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
  structure provides new insights into cellulosome assembly."
RL Structure 5:381-390(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Torno J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
  Steitz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
  domain: a general mechanism for attachment to cellulose."
RL EMBO J. 15:5739-5751(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402065;
RA Tavares G.A., Beguin P., Alzari P.M.;

```



```

FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).
FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).
FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).
FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).
FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).
FT ACT_SITE 162 162 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 338 338 T -> P (IN REF. 2).
FT CONFLICT 340 346 TPTTPT -> ROHOHQ (IN REF. 2).
SQ SEQUENCE 1331 AA: 146892 MW: 316851 BB8D8F0E0 CRC64;

Query Match 42.5%; Score 352; DB 1; Length 1331;
Best Local Similarity 45.8%; Pred. No. 6.7e-25;
Matches 71; Conservative 30; Mismatches 50; Indels 4; Gaps 3;

OY 3 GGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVRYWFTRDGSSFLVYNCDA 62
DB 568 GQIKVLYANKETNSTNTIRPMLKVVNSGSSSIDLSRVTIRWYTVVTDGERAQSALS-DWA 626

OY 63 AMGCNIRASFGSVNPATPTADYIQLSFTGCT--LAAGSGTGEIONRVKNSDWSNFDE 120
DB 627 QIGASNVTFKFKVLSSVSAGDYILEIGFKGAGQLQPKGTGEIQIREKNSDWSNYNOG 686

OY 121 NDYSY-CTNTAFQDMTKVTVYVNGRLVWGTPEPSGT 154
DB 687 NDWSIQSMTSYGENEKEVTAYIDGVLVWQEPSPGT 721

RESULT 9
GUNB_CALSA STANDARD; PRT: 1742 AA.
ID GUNB_CALSA
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'o V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum.";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: L32742; AAA91086.1; -.
DR EMBL: M36063; AAA72860.1; -.
DR EMBL: L01257; -. NOT_ANNOTATED_CDS.
DR PIR: A43745; A43745.
DR HSSP: P26221; 1TF4.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR000556; Glyco_hydro_48.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR Pfam: PF00942; CBM_3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR PRINTS: PR00844; GLHYDRLASE48.
DR PRODOM: P0001947; CBD_3; 2.
DR PRODOM: P0011903; Glyco_hydro_48; 1.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00598; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA: 193696 MW: 316899 A2123EED07 CRC64;

Query Match 42.5%; Score 352; DB 1; Length 1742;
Best Local Similarity 45.5%; Pred. No. 9.2e-25;
Matches 71; Conservative 30; Mismatches 51; Indels 4; Gaps 3;

OY 2 SGVKVQVKNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVRYWFTRDGSSFLVYNCDA 61
DB 704 SQGIKVLVANKETNSTNTIRPMLKVVNSGSSSIDLSRVTIRWYTVVTDGERAQSALS-DW 762

OY 62 AMGCNIRASFGSVNPATPTADYIQLSFTGCT--LAAGSGTGEIONRVKNSDWSNFDE 119
DB 763 AIGASNVTFKFKVLSSVSAGDYILEIGFKGAGQLQPKGTGEIQIREKNSDWSNYNO 822

OY 120 NDYSY-CTNTAFQDMTKVTVYVNGRLVWGTPEPSGT 154
DB 823 GNDWSIQSMTSYGENEKEVTAYIDGVLVWQEPSPGT 858

RESULT 10
GUNB_CALSA STANDARD; PRT: 1039 AA.
ID GUNB_CALSA
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellobiohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)
DE (1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
```


RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;
 RT "Nucleotide sequence of a gene from *Caldocellum saccharolyticum*
 RL encoding for exocellulase and endocellulase activity.";
 CC Nucleic Acids Res. 17:439-439(1989)
 CC -1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
 CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
 CC AN ENDOGLUCANASE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC F (FAMILY 10 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: X13602; CAA31936.1; .
 DR PIR: S02711; S02711.
 DR HSP: Q06851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001547; GH_5.
 DR InterPro: IPR001000; Glyco_hydro_10.
 DR Pfam: PF00150; cellulase; 1.
 DR Pfam: PF00331; Glyco_hydro_10; 1.
 DR Pfam: PF00942; CBM_3; 1.
 DR PRINTS: PR00134; GLHYDRLASE10.
 DR PRODOM: PD001947; CBD_3; 1.
 DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE: PS00659; Glycosylase; Repeat;
 KW Cellulose degradation; Hydrolase; Glycosidase;
 KW Multifunctional enzyme; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 1039 ENDOGLUCANASE/EXOGLUCANASE B.
 FT DOMAIN 376 416 THR/PRO-RICH, TANDM REPEATS OF T-P.
 FT DOMAIN 417 570 CELLULOSE-BINDING (BY SIMILARITY).
 FT DOMAIN 571 618 THR/PRO-RICH, TANDM REPEATS OF T-P.
 FT ACT_SITE 177 177 PROTON DONOR (POTENTIAL).
 FT ACT_SITE 285 285 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 792 792 BY SIMILARITY.
 FT ACT_SITE 1039 AA; 117641 MW; OE0378171594DDAE CRC64;
 SQ SEQUENCE 1039 AA; 117641 MW; OE0378171594DDAE CRC64;

Query Match 41.9%; Score 347; DB 1; Length 1039;
 Best Local Similarity 45.5%; Pred. No. 1.5e-24;
 Matches 70; Conservative 30; Mismatches 50; Indels 4; Gaps 3;

Qy 3 CGVKVQYKNNDSAPCDNOIKPGLQLVNTGSSSVLDSTVTYVWFTRDGSSTLVYNCDA 62
 Db 420 GOIKVYANKENTSTNTIRPWLKVVNGSSSIDLSRVTIRYKVTVDGSAQSAVS-DWA 478
 Qy 63 AMGCNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNDEET 120
 Db 479 QIGASNTVTFKVLSSSVSGADYLEICFKSGAGQLQPKGTGEIQIRFNKSDWSNRYNQG 538
 Qy 121 NDYSY-GTNTAFQDMTKVTVVYVNGRLVWGTEPSG 153
 Db 539 NDWSLQMSYSGENKVTAVIDGVLVNGQBPSPG 572

RESULT 11
 GUNW_ERWCA
 ID GUNW_ERWCA STANDARD; PRT; 504 AA.
 AC Q59395;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
 DE (Cellulase V1).
 GN CELV1.
 OS Erwinia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCC3193;
 RX MEDLINE=95231512; PubMed=7715600;
 RT Mae A., Heikinhelmo R., Palva E.T.;
 RT "Structure and regulation of the Erwinia carotovora subspecies
 RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
 RT phytopathogenicity.";
 RL Mol. Gen. Genet. 247:17-26(1995).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: X79241; CAA55823.1; .
 DR HSP: O85465; IA3H.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00150; cellulase; 1.
 DR Pfam: PF00942; CBM_3; 1.
 DR PRODOM: PD001947; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE: PS00659; Glycosylase; Signal.
 KW Cellulose degradation; Hydrolase; Glycosidase;
 FT SIGNAL 1 31
 FT CHAIN 32 504 ENDOGLUCANASE V1.
 FT DOMAIN 32 334 CATALYTIC.
 FT DOMAIN 335 352 LINKER.
 FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 41.4%; Score 343.5; DB 1; Length 504;
 Best Local Similarity 44.8%; Pred. No. 1.3e-24;
 Matches 69; Conservative 32; Mismatches 46; Indels 7; Gaps 4;

Qy 2 SGVKVQYKNNDSAPCDNOIKPGLQLVNTGSSSVLDSTVTYVWFTRDGSSTLVYNC 59
 Db 354 TGDIVLQYRVNVDNPNDDAIRMAFNKNTGSPPIKLSLDQVRYVPHDCKGPCANLFPV--- 410
 Qy 60 DWAAAGCCNIRASFGSVNPATPTADTYLQLSF-TGTTLAAGSGTGEIQNRVKNKSDWSNFD 118
 Db 411 DWANVGPNNTVTSTCTPAATDKANRYVLVTFASGSLQPCAGTGEVQVRIHAGDSNVN 470
 Qy 119 ETNDYSYCTN-TAFODTKVTVVYVNGRLVWGTEP 151
 Db 471 ETNDYSYGNITSYTNWTKITVHDKGTLVWGTEP 504

RESULT 12
 GUNW_ERWCA
 ID GUNW_ERWCA STANDARD; PRT; 505 AA.
 AC Q47096;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)


```

ID AC GUNZ_CLOS STANDARD; PRT; 986 AA.
AC P23659;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase 2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Thermoactive cellulase) (Avicelase I).
GN CELZ.
OS Clostridium stercorarium.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1510;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36 AND 475-486.
RC STRAIN=NCIB 11745;
RC MEDLINE=91066838; PubMed=2250652;
RX Jauris S., Ruecknagel K.P., Schwarz W.H., Kratzsch P.,
RA Bronnenmeier K., Staudenbauer W.L.;
RT "Sequence analysis of the Clostridium stercorarium celz gene encoding
a thermoactive cellulase (Avicelase I): Identification of catalytic
and cellulose-binding domains."
KL Mol. Gen. Genet. 223:258-267(1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -1- SIMILARITY: CONTAINS MULTIPLE CELLULOSE-BINDING SITES.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55299; CAA39010.1; ALT_SEQ.
DR PIR; S12021; S12021.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR005102; DUF291.
DR InterPro; IPR001701; GH_9.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF03442; DUF291; 2.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 986 ENDOGLUCANASE 2.
FT REPEAT 498 641 DOMAIN C'.
FT REPEAT 651 738 DOMAIN B'.
FT REPEAT 744 831 DOMAIN B'.
FT REPEAT 854 986 DOMAIN C.
FT DOMAIN 835 986 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 400 400 BY SIMILARITY.
FT ACT_SITE 438 438 BY SIMILARITY.
FT ACT_SITE 447 447 BY SIMILARITY.
SQ SEQUENCE 986 AA; 109512 MW; 1802E09B2292323690 CRC64;

Query Match 34.4%; Score 285; DB 1; Length 986;
Best Local Similarity 37.7%; Pred. No. 8,le-19;
Matches 58; Conservative 36; Mismatches 52; Indels 8; Gaps 3;

Qy 2 SGGVKVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSVTVTYVYWFTRDGGSTLVYNCW 61
Db 837 TGVLOIQMFNRSQDKTNGIPRLTNTGTPRLSDVKIRYYITDGEKDNQFW-CDW 895

Qy 62 AAMCGCNTRASFGSNPATPTADTYLQLSFTGCGTLAAG----GSGTEQNRVNKSQNSNF 117
Db 896 SSVGSNNTITGFVKMAEPKEGADYYLEFGTFD----AGYLPQNSIEVQNRFSKADWTDY 952
```

```

Qy 118 DETNDYSYGTNTAFQDWTKTVTYVNGKRLVNGTPEP 151
Db 953 IOTNDYSFSTNTSYGSNDRIITYISCVLVSGIEP 986

RESULT 15
GUNI_CLOTM
ID GUNI_CLOTM STANDARD; PRT; 879 AA.
AC 002934;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase)
DE (Cellulase I).
GN CELI.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.
RC STRAIN=NCIB 10682;
RX MEDLINE=93171873; PubMed=8436949;
RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
RA Gilbert H.J.;
RT "Gene sequence and properties of Celli, a family E endoglucanase from
Clostridium thermocellum."
RL J. Gen. Microbiol. 139:307-316(1993).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-HEXA-
GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04735; AAA20892.1;
DR PIR; A47704; A47704.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001701; GH_9.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR Pfam; PF00942; CBM_3; 2.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 55
FT CHAIN 56 879 ENDOGLUCANASE 1.
FT DOMAIN 56 518 CATALYTIC.
FT DOMAIN 729 879 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 448 448 BY SIMILARITY.
FT ACT_SITE 486 486 BY SIMILARITY.
FT ACT_SITE 495 495 BY SIMILARITY.
SQ SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;

Query Match 31.0%; Score 257; DB 1; Length 879;
Best Local Similarity 37.8%; Pred. No. 2,8e-16;
Matches 48; Conservative 34; Mismatches 41; Indels 4; Gaps 3;

Qy 1 VSGGVKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSVTVTYVYWFTRDGGSTLVYNC 60
Db 736 IKGEVLYQVANGNAGATSNIPFKIINNGTKAINLSVDKIRYYITKGGASQNF-CD 794
```



```
Db 79 QVGCNVRGCSFVKLTGRTGADYYIEITFTSGAGSLAAGASSGDIQVIRINKNDWTNYNEA 138
QY 121 NDYSYGT-TNTAFQDWTKVTVVYNGRLVWCTEP 151
DB 139 NDYSYDPTKTSFADNWRVTLYRNGQLWGVPE 170

RESULT 2
Q9L8L8
ID Q9L8L8 PRELIMINARY; PRT; 921 AA.
AC Q9L8L8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta-1,4-xylanase Xyna precursor.
GN Xyna.
OS Caldicellulosus cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Allicyclobacillaceae; Caldicellulosus.
OX NCBI_TaxID=74586;
KW SIGNAL.
FT CHAIN 1 33
SQ SEQUENCE 921 AA: 102380 MW: C5DDDIAT7F567413 CRC64;

Query Match 51.7%; Score 429; DB 2; Length 921;
Best Local Similarity 51.6%; Pred. No. 6.7e-27;
Matches 79; Conservative 35; Mismatches 37; Indels 2; Gaps 2;

QY 2 SGCVKQVQKNDSAPGDNQIKPGLQVLNTGSSVDLSTVTVRYWFTRDGSSTLVYNCDA 61
DB 563 SGTLRVEYRVGDSASATDNQMKPQLRIVNTGSAVPLTELKVRWYTKN-STQAEQYFCDW 621

QY 62 AAMGCGNIRASFGSNPATPTADTYLQLSFTGTLAAGSGTGEIONRVNKSWSNFDET 121
DB 622 AQICGSNIRAFVSLAQPVSGADSVIELSFTGSGVPAGNTEIONRHFTHNMNINETD 681

QY 122 DYSY-GTNTAFQDWTKVTVVYNGRLVWCTEP 153
DB 682 DWSYNGTOTTWGPSTRITLYRNGVLVWCTEP 714

RESULT 3
Q9RFX5
ID Q9RFX5 PRELIMINARY; PRT; 930 AA.
AC Q9RFX5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN MANA.
OS Caldicellulosus cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
```

```
OC Allicyclobacillaceae; Caldicellulosus.
OX NCBI_TaxID=74586;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosus cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL: AF163837; AAF22274.1; -.
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR004302; Chitin_binding_3.
DR InterPro: IPR001547; GH_5.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF00942; CBM_3; 2.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF03067; Chitin_bind_3; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR ProDom: PD001947; CHD_3; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW SIGNAL.
FT CHAIN 1 33
SQ SEQUENCE 930 AA: 101576 MW: 0086638D54D1A2CC CRC64;

Query Match 51.7%; Score 429; DB 2; Length 930;
Best Local Similarity 51.3%; Pred. No. 6.7e-27;
Matches 78; Conservative 31; Mismatches 39; Indels 4; Gaps 3;

QY 3 GGVKQVQKNDSAPGDNQIKPGLQVLNTGSSVDLSTVTVRYWFTRDGSSTLVYNCDA 62
DB 780 GNLVQVYRAADTNATDNQKLPFRIVNCRCTSSVPLSELTYWYTVYD-GDKQVQVNCDA 838

QY 63 AAMGCGNIRASFGSNPATPTADTYLQLSFT--GGTLAAGSGTGEIONRVNKSWSNFDET 120
DB 839 QVGCNLRGCSFVKLTGRTGADYYIEITFTSGAGSLAAGASSGDIQVIRINKNDWTNYNEA 898

QY 121 NDYSYGT-TNTAFQDWTKVTVVYNGRLVWCTEP 151
DB 899 NDYSYDPTKTSFADNWRVTLYRNGQLVWCTEP 930

RESULT 4
Q9Z4I1
ID Q9Z4I1 PRELIMINARY; PRT; 997 AA.
AC Q9Z4I1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cellulase precursor (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=89769;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL: AJ133614; CAB38941.1; -.
DR HSSP: P26221; 1TF4.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR000566; Lipocln_cytFABP.
```

DR	Pfam:	PF00942;	CBM_3; 2.		
DR	Pfam:	PF00041;	fn3; 2.		
DR	Pfam:	PF00759;	Glyco_hydro_9; 1.		
DR	ProDom:	PD001947;	CBD_3; 1.		
DR	SMART:	SMO0060;	FN3; 2.		
DR	PROSITE:	PS00592;	GLYCOSYL_HYDROL_F9_1; 1.		
DR	PROSITE:	PS00698;	GLYCOSYL_HYDROL_F9_2; 1.		
DR	PROSITE:	PS00213;	LIPOCALIN; UNKNOWNW_1.		
KW	Glycosidase; Hydrolase; Signal.				
FT	SIGNAL	1	34		
FT	POTENTIAL.				
SEQ	SEQUENCE	997 AA;	106927 MW; F20CB119D6410629 CRC64;		
<hr/>					
Query Match 46.4% Score 385; DB 2; Length 997;					
Best Local Similarity 48.4%; Pred. No. 2.9e-23;					
Matches 74; Conservative 29; Mismatches 46; Indels 4; Gaps 3;					
<hr/>					
Qy	2	SGGVKVOYKNNDSSAPGDNQIKPGQLVNTGSSVDLSTVTVRYWFTRDGGSTLVYNCDW	61		
Db	846	TCTLEVOYRSCGSNNAVPQFNLKNTQTQAIDLSTVKIRYFTKD-GTEELSFWCXY	904		
<hr/>					
Qy	62	AAMCGGNTRASFGSVNPATPTADTYLQLSFT--CGTLAAGSTGEIQNRVNKSOWSNDFE	119		
Db	905	AVGSAANVGQMEFAVNPAAKTADTYVEISFTSGAGSLAAGAETGYIOTRFSKNNWSAFDQ	964		
<hr/>					
Qy	120	TNDYSY-CGTNTAFODWTKVTVVYNGRLVMGTEP	151		
Db	965	SDNYSDASKTAFPAANKVTAQGNTQVMGLEP	997		
<hr/>					
RESULT	5				
Q93TJ6	PRELIMINARY; PRT; 499 AA.				
ID	Q93TJ6				
AC	O93TJ6;				
DT	01-DEC-2001 (TrEMBLrel. 19, Created)				
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)				
DE	Alkali tolerable cellulase.				
GN	CEL.				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;				
OC	Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Y106;				
RA	Qi M., Wang P., Liu X., Qu Y.;				
RT	"The cloning and expression of a cellulase from Bacillus subtilis				
RT	YT106.";				
RL	Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.				
DR	EMBL:	AF35629; AAK3940.1; -			
DR	InterPro:	IPR001956; CBD_3.			
DR	InterPro:	IPR001547; GH_5.			
DR	Pfam:	PF00942; CBM_3; 1.			
DR	Pfam:	PF00150; cellulase; 1.			
DR	ProDom:	PD001947; CHD_3; 1.			
DR	PROSITE:	PS00659; GLYCOSYL_HYDROL_F5; UNKNOWNW_1.			
SEQ	SEQUENCE	499 AA;	55252 MW; 587A687584B7E5BE CRC64;		
<hr/>					
Query Match 44.6% Score 370; DB 2; Length 499;					
Best Local Similarity 44.0%; Pred. No. 2.1e-22;					
Matches 66; Conservative 34; Mismatches 46; Indels 4; Gaps 2;					
<hr/>					
Qy	4	GVKVOYKNNDSSAPGDNQIKPGQLVNTGSSVDLSTVTVRYWF--TROGGSTLVYNCDA	62		
Db	353	GISVOYRAGDSGMNSNOIRPOLQTKNGNTVDLDKYTRYWKAKNKQN---FDGDA	409		
<hr/>					
Qy	63	AAMCGGNTRASFGSVNPATPTADTYLQLSFTGGTLAAGSTGEIQNRVNKSOWSNDFETND	122		
Db	410	QIGCGNVTHKFVLHLPKQADITYLELGFKNGLAPCASTGNLTQLRLHNDDWSNVAQSGD	469		
<hr/>					
Qy	123	YSYGNTAFODWTKVTVVYNGRLVMGTEPS	152		
Db	470	YSFFKSNFTFKTKITLYDOCKLINGTEPN	499		

```
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00942; CBM_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 499 AA: 55193 MW; 990A1878788CF738 CRC64;

Query Match 43.8%; Score 363; DB 2; Length 499;
Best Local Similarity 43.38; Pred. No. 8e-22;
Matches 65; Conservative 33; Mismatches 48; Indels 4; Gaps 2;

Qy 4 GVKVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSSTLVYNCOWA 62
Db 353 GTSVOYRAGDGSMSNSQIRPOLIKNNNGNTVDLKDVTARYWYKAKNKGQNV---DCDYA 409

Qy 63 AMGCNIRASFGSVNPTPTADTYLQLSFTGCTLAAGSGTGEIQNRVYKNSDWSNFDND 122
Dy 410 QIGCGNLTHKEVTLHKPKQGADTYLELFGKNTLPGPGASTGNIQLRLHNDWSNYAOSGD 469

Qy 123 YSYGTNTAFQDWTKVTVVYVNGRLVWGTEPS 152
Dy 470 YSFFQSNTEFTTKKITLYHQGLKINGTEPN 499

RESULT 8
Q45532 PRELIMINARY; PRT; 499 AA.
AC Q45532;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cellulase.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87190397; PubMed=3106035;
RA Nakamura A., Uozumi T., Beppu T.;
RT "Nucleotide sequence of a cellulase gene of Bacillus subtilis.";
RL Eur. J. Biochem. 164:317-320(1987).
DR EMBL: M28332; AAA22307.1; -.
DR HSSP: O85465; 1A3H.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00942; CBM_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 499 AA: 55075 MW; D424AB9E53B94D23 CRC64;

Query Match 43.4%; Score 360; DB 2; Length 499;
Best Local Similarity 43.68; Pred. No. 1.4e-21;
Matches 65; Conservative 32; Mismatches 48; Indels 4; Gaps 2;

Qy 4 GVKVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSSTLVYNCOWA 62
Dy 353 GTSVOYRAGDGSMSNSQIRPOLIKNNNGNTVDLKDVTARYWYKAKNKGQNV---PCDYA 409

Qy 63 AMGCNIRASFGSVNPTPTADTYLQLSFTGCTLAAGSGTGEIQNRVYKNSDWSNFDND 122
Dy 410 QIGCGNLTHKEVTLHKPKQGADTYLELFGKNTLPGPGASTGNIQLRLHNDWSNYAOSGD 469

Qy 123 YSYGTNTAFQDWTKVTVVYVNGRLVWGTEP 151
Dy 470 YSFFQSNTEFTTKKITLYHQGLKINGTEP 498

RESULT 9
P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Endoglucanase A (EC 3.2.1.4) (Endo-1,4-beta-glucanase A) (Cellulase A) (Fragment).
DE CELA.
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Anaerocellum group; Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=2-1320;
RA Zverlov V.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -!- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULOSE FAMILY E
CC (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULOSE FAMILY L
CC (FAMILY 48 OF GLYCOSYL HYDROLASES).
DR EMBL: Z86105; CAB06786.1; -.
DR HSSP: P26221; 1TF4.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR000556; Glyco_hydro_48.
DR Pfam: PF00942; CBM_3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS: PR00844; GLHYDRLASE48.
DR ProDom: PD001947; CBD_3; 2.
DR ProDom: PD011903; Glyco_hydro_48; 1.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT NON_TER 1
SQ SEQUENCE 1711 AA: 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 43.4%; Score 359.5; DB 2; Length 1711;
Best Local Similarity 46.68; Pred. No. 6.6e-21;
Matches 76; Conservative 25; Mismatches 45; Indels 17; Gaps 5;

Qy 1 VSGG-VKVOYKNDSPAGDNOIKPGLQIVNTGSSVDLSTVTYRYWFTHDGSSSTLVYNC 59
Dy 882 VAGGQIKVLYANKETNTTIRPWLKVVNTGSSSIDLSRVTIRYWTVDGDKAUSAIS- 940

Qy 60 DWAAMGCCNIRASFGSVNPTPTADTYLQLSFTGCT--LAAGSGTGEIQNRVYKNSDWSNF 117
Dy 941 DWAQIGASNVTFKPVKLVSSVSGADYILEIGFKSGAGOLQAGKDTGEIQIRPNKSDWSNY 1000

Qy 118 DETNDYS-----YGTNTAFQDWTKVTVVYVNGRLVWGTEPSG 153
Dy 1001 NQGDWSHWQSWNTNYGEN-----VKVTAYIDGVLVWGTEPSG 1037

RESULT 10
O83012 PRELIMINARY; PRT; 501 AA.
AC O83012;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cellulase.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
```



```

RN  [1]
RP  SEQUENCE FROM N.A.
RA  Khandongnuch C., Ooi T., Kinoshita S.;
RT  "Cloning and nucleotide sequence of beta-mannanase and cellulase gene
RT  from Bacillus sp. 5H.";
RL  Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AB016164; BAA31712.1; -.
DR  HSSP; O85465; 1A3H.
DR  InterPro; IPR001956; CBD_3.
DR  InterPro; IPR001547; GH_5.
DR  Pfam; PF00942; CBM_3; 1.
DR  Pfam; PF00150; cellulase; 1.
DR  PRINTS; PR01652; SHAPEPROTEIN.
DR  ProDom; PD001947; CBD_3; 1.
DR  PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ  SEQUENCE 501 AA; 55422 MW; F76D260A901E2D1D CRC64;

Query Match 42.8%; Score 355; DB 2; Length 501;
Best Local Similarity 43.0%; Pred. No. 3.6e-21;
Matches 64; Conservative 33; Mismatches 48; Indels 4; Gaps 2;

QY  4 GVKVQYKNNDSAPGDNOIKPGLQLVNTGSSVDLSVTYVRYWFTTRDGGSTLVYNCDA 62
DB  1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:|||||
353 GISVQTKAGDGGVNSNOIRPOLHIKNGNATVDLKVDTARYWYNAKNGQN---FDCDYA 409
QY  63 AMCGCGRIRASFGSVNPATPTADTYLQLSFTGCT--LAAGCGSTGEIQNRVYKSDWSNFDETND 122
DB  1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:|||||
410 QICCGNLTHKFTVLHKPKGADTYLELGFKKGLSPCASTGNTQLRLHNDWSNYAQSGD 469
QY  123 YSGTNTAFODWTKVTVVYVNGRLVWGTEP 151
DB  1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:|
470 YSPFSQNTFTKTKITLYNHGKLIWGCTEP 498

RESULT 11
O24820 PRELIMINARY; PRT: 1000 AA.
AC O24820;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
RT thermophilic anaerobe strain NA10.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008029; BAA22939.1; -.
DR HSSP; O06851; INBC.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001547; GH_5.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 42.7%; Score 354; DB 2; Length 1000;
Best Local Similarity 45.8%; Pred. No. 9.9e-21;
Matches 71; Conservative 30; Mismatches 50; Indels 4; Gaps 3;

```

```

QY  2 SGGVKVQYKNNDSAPGDNOIKPGLQLVNTGSSVDLSVTYVRYWFTTRDGGSTLVYNCDA 61
DB  1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:|||||
372 SGOIKVLANKETSTNTIRPWLKVVNTGSSVDLSRVTYWYVTDGDKAQSASV-DW 430
QY  62 AMCGCGRIRASFGSVNPATPTADTYLQLSFTGCT--LAAGCGSTGEIQNRVYKSDWSNFDE 119
DB  1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:|
431 AQIGASNVTFKFKVLKSSVSGADYYLEIGFKSGAGQLQPGKDTGEIQIRFNKSDWSNYNQ 490
QY  120 TNDYSY-CTNTAFODWTKVTVVYVNGRLVWGTEPSG 153
DB  1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:|
491 GNDMSWTSQSMSTSYGENMKVTAIDGVLVWGQEPG 525

RESULT 12
Q9A0G4 PRELIMINARY; PRT: 1751 AA.
AC Q9A0G4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cele.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Syntrophomonadaceae; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TOK7B.1;
RX MEDLINE-20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001701; GH_9.
DR InterPro; IPR000566; Lipocln_cytFABP.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF00942; CBM_3; 4.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR ProDom; PD001947; CBD_3; 3.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 42.0%; Score 348; DB 2; Length 1751;
Best Local Similarity 44.7%; Pred. No. 5.9e-20;
Matches 71; Conservative 25; Mismatches 47; Indels 16; Gaps 4;

QY  4 GVKVQYKNNDSAPGDNOIKPGLQLVNTGSSVDLSVTYVRYWFTTRDGGSTLVYNCDA 63
DB  1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:|||||
678 GVKVLKNNETSASTGIRPWEKIVGSSVDLSRVKIRYWYVTDGDKPOSASV-CDMAQ 736
QY  64 MCCGCRIRASFGSVNPATPTADTYLQLSFTGCT--LAAGCGSTGEIQNRVYKSDWSNFDET 121
DB  1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:|
737 IGASNVTFNFVKLSSVSGADYYLEVGFSGAGLOLPGKDTGDIQVRFNKNDWSNYNQAD 796
QY  122 DYS-----YGTNTAFODWTKVTVVYVNGRLVWGTEPSG 153
DB  1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:| 1:|:|:|
797 DWSLQSMYNGEN-----AKVTLYVDGLVWGQEPG 829

RESULT 13
Q9X3P5 PRELIMINARY; PRT: 1770 AA.
AC Q9X3P5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

```


1172 1172 ICASNVT'FNFVKLTSGVSGADYYLEVGFSSGACQLQPGKDT'GDIQVR'FNKNDWSNYNQAD 1231

Qy 122 DYS-----YGTNTAFQDWTKVTVVYVNGRLVWGTEPSG 153

Ub 1232 DWSWMQSMNTNYGEN-----AKVTL'YVDGLVWVGQEPGG 1264

Search completed: November 13, 2002, 11:54:11
Job time : 16.2716 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: November 13, 2002, 11:49:46 : Search time 4.75718 Seconds
(without alignments)
952.482 Million cell updates/sec

Title: US-09-917-378-4

Perfect score: 829

Sequence: 1 VSGGVKKVQYKNNDSAPGDNQ.....TKVTVVYVNGRLVWMEPSGT 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA*
1: /cgn2_6/ptodata/1/laa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/laa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/laa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/laa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/laa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/laa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374	45.1	551	2	US-09-033-537A-1
2	373.5	45.1	700	2	US-07-862-588B-2
3	357.5	43.1	493	4	US-09-198-956-10
4	357.5	43.1	493	4	US-09-198-955A-12
5	357.5	43.1	493	4	US-09-694-531-12
6	357.5	43.1	493	4	US-09-670-141-10
7	356.5	43.0	167	5	PCT-US95-13813-9
8	349	42.1	616	4	US-09-136-574A-47
9	348	42.0	1751	4	US-09-136-574A-44
10	340	41.0	1426	4	US-09-136-574A-43
11	226	27.3	162	1	US-08-048-164A-2
12	226	27.3	162	1	US-08-460-462-2
13	226	27.3	162	1	US-08-460-457-2
14	226	27.3	162	1	US-08-460-458-2
15	226	27.3	162	2	US-08-460-455-2
16	226	27.3	162	2	US-08-330-394A-2
17	226	27.3	163	3	US-09-006-636-7
18	226	27.3	163	4	US-09-006-632-7
19	226	27.3	163	4	US-09-325-274-7
20	225	27.1	382	4	US-09-277-716-22
21	225	27.1	382	4	US-09-609-161B-22
22	224	27.0	156	2	US-08-330-394A-22
23	218	26.3	154	2	US-08-330-394A-29
24	127	15.3	531	2	US-07-862-588B-7
25	85	10.3	428	3	US-09-118-319-5
26	85	10.3	2123	4	US-08-968-685A-10
27	84	10.1	434	2	US-09-008-962-1

Query Match 45.1% Score 374; DB 2; Length 551;

ALIGNMENTS

RESULT 1

US-09-033-537A-1
: Sequence 1, Application US/09033537A
: Patent No. 5958083

: GENERAL INFORMATION:
: APPLICANT: Onishi, Masahiro

: APPLICANT: Fich, Merette
: APPLICANT: Toft, Annette Hanne

: APPLICANT: Sh lein, Martin
: TITLE OF INVENTION: Prevention Of Back-Staining

: TITLE OF INVENTION: In Stone Washing
: NUMBER OF SEQUENCES: 1

: CORRESPONDENCE ADDRESS:

: ADDRESSER: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
: STREET: 405 Lexington Avenue
: CITY: New York

: STATE: NY

: COUNTRY: U.S.A.

: ZIP: 10174

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: DOS

: SOFTWARE: FastSEQ for Windows Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/033.537A

: FILING DATE: 02-MAR-1998

: CLASSIFICATION: 008

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 0993/95

: FILING DATE: 08-SEP-1995

: APPLICATION NUMBER: PCT/DK96/00364

: FILING DATE: 03-SEP-1996

: ATTORNEY/AGENT INFORMATION:

: NAME: Green, Reza

: REGISTRATION NUMBER: 38,475

: REFERENCE/DOCKET NUMBER: 4492.204-US

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 212-867-0123

: TELEFAX: 212-878-9655

: TELEX:

: INFORMATION FOR SEQ ID NO: 1:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 551 amino acids

: TYPE: amino acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: US-09-033-537A-1

Sequence 1, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 48, Appli
Sequence 23, Appli
Sequence 23, Appli
Sequence 9, Appli
Sequence 9, Appli
Sequence 10, Appli
Sequence 10, Appli
Sequence 9, Appli
Sequence 10, Appli
Sequence 3, Appli
Sequence 5, Appli

28 84 10.1 434 2 US-08-675-507-1
29 84 10.1 434 3 US-09-213-205-1
30 84 10.1 1833 4 US-08-621-944A-4
31 84 10.1 1833 4 US-08-945-567D-4
32 84 10.1 1833 4 US-08-621-944A-3
33 84 10.1 1992 4 US-08-945-567D-3
34 82 9.9 2048 4 US-09-268-347-48
35 80 9.7 326 4 US-09-286-691-23
36 80 9.7 326 4 US-09-687-147-23
37 80 9.7 1338 2 US-08-728-470-9
38 80 9.7 1338 4 US-08-719-641-9
39 80 9.7 1529 2 US-08-728-470-10
40 80 9.7 1529 4 US-08-719-641-10
41 80 9.7 1599 2 US-08-617-697-9
42 80 9.7 1600 2 US-08-617-697-10
43 79.5 9.6 333 3 US-08-988-111-3
44 79.5 9.6 333 4 US-09-387-922-3
45 79.5 9.6 892 4 US-09-336-447A-5

```

Rest Local Similarity   47.18;  Pred. No. 8.le-30;
Matches    72;  Conservative      33; Mismatches    42; Indels      6; Gaps      4

Oy     2 SGGVKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGSSTLVYNCDW 61
       : : ||||| : ||||| : | : : ||||| : ||||| : ||||| : ||||| : ||
Db     402 TGNLVQYKVCGDTSATDNQMRPSNKKNGTTPVNI.SGLK.I.HYYFTKD-GTADMSASFDM 460
Oy     62 AAMCGCNIRASFGSVNPATPTADTYVLQSFT--GGTLAAGSGTGEIQNKVNKSDFNSFDE 119
       : : || : || : | : ||||| : ||||| : ||||| : ||||| : ||||| : ||
Db     461 AQIGASNVSAAF--ANFTGTSNTDITVELLSFSAGSGSIPAGCQTGDIQLRMKYTKDSNFNE 518
Oy     120 TNDYSY-CTNTAFODWTKTVTYVYVNCRLVMGTETP 151
       ||||| : || : || : ||||| : ||||| : ||||| : ||||| : ||||| : ||
Db     519 ANDSYDGAKTAYADNRRTYLHQNLCTLVMGTTTP 551

RESULT 2
US-07-862-588B-2
: Sequence 2, Application US/07862588B
: Patent No. 5916796
: GENERAL INFORMATION:
: APPLICANT: Joergensen, Per Linnaa
: APPLICANT: Sch Ieln, Martin
: APPLICANT: Hansen, Christian
: TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESSES:
: ADDRESSSEE: NO. 59167960 No. 5916796disk of No. 5916796th America, Inc.
: STREET: 405 Lexington Avenue, 62nd floor
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10017
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/862.588B
: FILING DATE: 19920727
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK 164/90
: FILING DATE: 19-JAN-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/DK91/00013
: FILING DATE: 18-JAN-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Zetison, Steve T./ Lambiris, Elias J.
: REGISTRATION NUMBER: 30,335 / 33,728
: REFERENCE/DOCKET NUMBER: 3425.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212 867 0123
: TELEFAX: 212 867 0298
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 700 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-07-862-588B-2

Query Match          45.1%; Score 373.5; DB 2; Length 700;
Best Local Similarity 49.7%; Pred. No. 1.3e-29;
Matches    77;  Conservative      32; Mismatches    39; Indels      7; Gaps      5

Oy     1 VSGGVKVOYKKNDD-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGSSTLVYNVC 59
       : : ||||| : | : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||
Db     549 VNSDLVVQYKDGDRNNATDNOIKPHENIQNKGTSPDLSSLTLRYYFTKD-SSAAANGWI 607
Oy     60 DWAAMCCGNIRASFGSVNPATPTADTYLQLSFT--GGTLAAGSGTGEIQNRVNKSDWSNF 117
       || : || : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
```

```

Query Match          43.1%: Score 357.5; DB 4; Length 493;
Best Local Similarity 42.3%; Pred. No. 3.3e-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY      1 VSGGVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTIVRYWFTRDGGSSTLVYNCD 60
       111 :|||: ||: : | | : : |||::||| :|:||: || : ||
Db     331 VSGNLKVFEFNSPNSDTTNSINPOFKVTNTGSSAIDLKLRLYYTYTVDGOKDQTFW-CD 389
       111 :|||: ||: : | | : : |||::||| :|:||: || : ||

QY      61 WAAM-----CGCNGIRASFSGVNPAITPATDTYLQLSFTGGTLAAGSGTGCEIQNKVK 111
       111 :||: :||: :| :| :| :|||::||| ||| :||| :|
Db     390 HAAIIGNSGYNGITSNVKGTFFVKMSNSTNADTYLEISFTGCTLEPGAHV-OIOCRFAK 448
       111 :||: :||: :| :| :| :|||::||| ||| :||| :|

QY     112 SDWSNFDETNDYSYGTTATFODWTKTVTIVYVNGRLVMWGTEPSGT 154
       :||||: ||||: : : | :| :||| ||| ||| :||:
Db    449 NDWSNYTQSNDYSFKRSQFVEWDQVTAYLNGVLVMGKEPGCS 491
       :||||: ||||: : : | :| :||| ||| ||| :||:

RESULT 6
US-09-670-141-10
: Sequence 10, Application US/09670141
: Patent No. 6429000
: GENERAL INFORMATION:
: APPLICANT: Andersen, Lene N.
: APPLICANT: Schulein, Martin
: APPLICANT: Lange, Niels Erik K.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Schnorr, Kirk
: TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
: FILE REFERENCE: Licheniformis
: CURRENT APPLICATION NUMBER: 5377.200-US
: CURRENT FILING DATE: 2000-09-26
: PRIOR APPLICATION NUMBER: 09/198,956
: PRIOR FILING DATE: 1998-11-24
: PRIOR APPLICATION NUMBER: 1344/97
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/067,240
: PRIOR FILING DATE: 1997-12-02
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 10
: LENGTH: 493
: TYPE: PRT
: ORGANISM: Bacillus licheniformis
US-09-670-141-10

Query Match          43.1%: Score 357.5; DB 4; Length 493;
Best Local Similarity 42.3%; Pred. No. 3.3e-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;

QY      1 VSGGVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTIVRYWFTRDGGSSTLVYNCD 60
       111 :|||: ||: : | | : : |||::||| :|:||: || : ||
Db     331 VSGNLKVFEFNSPNSDTTNSINPOFKVTNTGSSAIDLKLRLYYTYTVDGOKDQTFW-CD 389
       111 :|||: ||: : | | : : |||::||| :|:||: || : ||

QY      61 WAAM-----CGCNGIRASFSGVNPAITPATDTYLQLSFTGGTLAAGSGTGCEIQNKVK 111
       111 :||: :||: :| :| :| :|||::||| ||| :||| :|
Db     390 HAAIIGNSGYNGITSNVKGTFFVKMSNSTNADTYLEISFTGCTLEPGAHV-OIOCRFAK 448
       111 :||: :||: :| :| :| :|||::||| ||| :||| :|

QY     112 SDWSNFDETNDYSYGTTATFODWTKTVTIVYVNGRLVMWGTEPSGT 154
       :||||: ||||: : : | :| :||| ||| ||| :||:
Db    449 NDWSNYTQSNDYSFKRSQFVEWDQVTAYLNGVLVMGKEPGCS 491
       :||||: ||||: : : | :| :||| ||| ||| :||:

RESULT 7
PCT-US95-13813-9
: Sequence 9, Application PC/TUS9513813
: GENERAL INFORMATION:
: APPLICANT: Yeda Research and Development Co. Ltd.
: APPLICANT: Ramot University Authority for Applied
: APPLICANT: Research and Industrial Development Ltd.
: APPLICANT: Technion Research and Development Foundation Ltd.
: APPLICANT: Bayer, Edward A.
: APPLICANT: Morag, Ely

```

APPLICANT: Wilchek, Meir
APPLICANT: Lamed, Raphael
APPLICANT: Shoham, Yuval
TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN ("BD")
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT NUMBER: PCT/US95/13813
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: BAYER-3
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 43.0%; Score 356.5; DB 5; Length 167;
Best Local Similarity 42.3%; Pred. No. 1e-28;
Matches 69; Conservative 36; Mismatches 47; Indels 11; Gaps 3;
QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQVLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYVNC 60
DB 5 VSGNLKVEFYNSPDTTNSINPQFKVTVNTGSSAIDLSKLTLLRYVTVYVVDGQKDTQFW-CD 63

QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSGTGEIONRVNK 111
DB 64 HAAIIGNSGYSNGITSNVRGTPYKMSSTNNATYLEISFTGCTLEPGARV-QIOGRFAK 122
QY 112 SDMSNFDENSYGNTAFQDMTKVTYVYVNGRLVWVGTEPSGT 154
DB 123 NDMSNYTQSDNYSFKXSQFVENDQVYAVINGVLVWVGTEPGCS 165

RESULT 8
US-09-136-574A-47
Sequence 47, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
ZIP: 20004

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 42.1%; Score 349; DB 4; Length 616;
Best Local Similarity 43.8%; Pred. No. 3.2e-27;
Matches 71; Conservative 26; Mismatches 49; Indels 16; Gaps 4;

QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQVLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYVNC 60
DB 1 MCGSVKLVYKNNETSASTGSIRPMFKLVNCGSSVDLSRVKIRYWTVDGDKPQSAV-CD 54
QY 61 WAAMCGCNTRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIONRVNKSDMSNFD 118
DB 60 WAQIGASNTVFNKLVKSSGVSGADYILEVGFSSGACQLQPKGKDTGDIQVRFNKNDMSNYN 119
QY 119 ETNDYS-----YGTNTAFQDMTKVTYVYVNGRLVWVGTEPSG 153
DB 120 QADDMSWLSQMTNYGEN-----AKVTLYVDGVLVWVGTEPGC 155

RESULT 9
US-09-136-574A-44
Sequence 44, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA


```
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136.574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1751 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

Query Match 42.0%; Score 348; DB 4; Length 1751;
Best Local Similarity 44.7%; Pred. No. 1.6e-26;
Matches 71; Conservative 25; Mismatches 47; Indels 16; Gaps 4;

Qy 4 GVKVOYKNDAPGDNQIKPGQLQLVNTGSSVVDLSTVTVRVWFTRDGSGSSTLVVNCWAA 63
Db 678 GVKVLYKNNETASGTSIRPFKIVNGSSVVDLSRVKIRVWYTVDDGKPOSAY-CDWAA 736

Qy 64 MCGGNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFDFTN 121
Db 737 IGASNVTFNFVKLSGSGVADYILEVGFSSGAGLOPKDGTGDIQVRFNKNKDSNYNQAD 796

Qy 122 DYS-----YGTNTAFQDWTKVTVVYVNGRLVWGTEPSG 153
Db 797 DWSWLQSMNTYGEN-----AKVTLYVDGVLVWGQEPGG 829

RESULT 10
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; ANDERSON, Paige
; GIBBS, Moreland
; BERGQUIST, Peter
; DANIELS, Roy
; MORGAN, Hugh W.
; WILLIAMS, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136.574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6294366e
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43

Query Match 41.0%; Score 340; DB 4; Length 1426;
Best Local Similarity 43.4%; Pred. No. 7.8e-26;
Matches 69; Conservative 26; Mismatches 48; Indels 16; Gaps 4;

Qy 4 GVKVOYKNDAPGDNQIKPGQLQLVNTGSSVVDLSTVTVRVWFTRDGSGSSTLVVNCWAA 63
Db 413 GLKVLVYKNNETASGTSIRPFKIVNGSSVVDLSRVKIRVWYTVDDGKPOSAY-CDWAA 471

Qy 64 MCGGNIRASFGSVNPATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDWSNFDFTN 121
Db 472 IGASNVTFNFVKLSGSGVADYILEVGFSSGAGLOPKDGTGDIQVRFNKNKDSNYNQAD 511

Qy 122 DYS-----YGTNTAFQDWTKVTVVYVNGRLVWGTEPSG 153
Db 532 DWSWLQSMNTYGEN-----AKVTLYVDGVLVWGQEPGG 564

RESULT 11
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shplegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
```

```

:      TOPOLOGY: linear
:      MOLECULE TYPE: protein
US-08-460-462-2

Query Match      27.3%  Score 226;  DB 1;  Length 162;
Best Local Similarity 34.8%;  Pred. No. 1.5e-15;
Matches 57;  Conservative 31;  Mismatches 58;  Indels 18;  Gaps 7;

QY  2  SGGVKVQYKKNDSAPGDNOIKPGLQLVNTGSSVDLSITVTYRYWFTTRDGGSTLVYNCIDW 61
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB  3  TSSMSVEFYNSKSAQNTSITPIIKITNTSDSLNLNDVKRYYYTSDGTGCTGCTFW-CDH 61
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  62  AAMCGN-----IRASF--GSVNPATPTADTYLQLSFTGGTLAAGCSTGE---IQNK 108
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB  62  AGALLGNSYVDNTSKVTANFVKETASIP-TSPYDIYVHFQFASG--AATLKKGQFTTQGR 118
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  109  VNKSDWSNFDFTNDYSYGTNTAFQDWT-KVTIVYVNGRIYVWCTEP 151
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB  119  ITRKDSWSNYTOTNDYSFDAQSSSTPVNPKVTGYICGAKVLGTAP 162
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 13
US-08-460-457-2
:  Sequence 2, Application US/08460457
:  Patent No. 5719044
:  GENERAL INFORMATION:
:  APPLICANT: Shoseyov, Oded
:  APPLICANT: Shpiegl, Itai
:  APPLICANT: Goldstein, Marc A.
:  APPLICANT: Doi, Roy H.
:  TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
:  NUMBER OF SEQUENCES: 21
:  CORRESPONDENCE ADDRESS:
:  ADDRESSEE: PENNIE & EDMONDS
:  STREET: 1155 Avenue of the Americas
:  CITY: New York
:  STATE: New York
:  COUNTRY: U.S.A.
:  ZIP: 10036
:  COMPUTER READABLE FORM:
:  MEDIUM TYPE: Floppy disk
:  COMPUTER: IBM PC compatible
:  OPERATING SYSTEM: PC-DOS/MS-DOS
:  SOFTWARE: PatentIn Release #1.0, Version #1.25
:  CURRENT APPLICATION DATA:
:  APPLICATION NUMBER: US/08/460.457
:  FILING DATE: Concurrently herewith
:  CLASSIFICATION: 435
:  PRIOR APPLICATION DATA:
:  APPLICATION NUMBER: US 08/048,164
:  FILING DATE: 14-APR-1993
:  ATTORNEY/AGENT INFORMATION:
:  NAME: Mirock, S. Leslie
:  REGISTRATION NUMBER: 18,872
:  REFERENCE/DOCKET NUMBER: 7809-008
:  TELECOMMUNICATION INFORMATION:
:  TELEPHONE: (212) 790-9090
:  TELEFAX: (212) 869-8864/9741
:  TELEX: 66141 PENNIE
:  INFORMATION FOR SEQ ID NO: 2:
:  SEQUENCE CHARACTERISTICS:
:  LENGTH: 162 amino acids
:  TYPE: amino acid
:  TOPOLOGY: linear
:  MOLECULE TYPE: protein
US-08-460-457-2

Query Match      27.3%  Score 226;  DB 1;  Length 162;
Best Local Similarity 34.8%;  Pred. No. 1.5e-15;
Matches 57;  Conservative 31;  Mismatches 58;  Indels 18;  Gaps 7;

QY  2  SGGVKVQYKKNDSAPGDNOIKPGLQLVNTGSSVDLSITVTYRYWFTTRDGGSTLVYNCIDW 61

```

Db	3	TSSMSVEFYNSNSAQTNSITPILIKITNTSDSLNDVKRVYYTSDGTQGTWF-CDII	61
QY	62	ANMCCGN-----IRASP--CSVNPATPATUTYQLQLSFTGCTLAAGSTGE---	IQNR 108
Db	62	AGALLGNSYVDNTSKVTANFKETASP-TSTYDTYVEFGFASG--AATLKQGFITIQGR	118
QY	109	VNKSDMNSFDETNDYSGYNTAFQDWT-KVTYVYNGRLVWGTEP	151
Db	119	ITKSDMSNYTQINDYGFDSADSSPPVNVNPKVTGYIGGAKVLGTAP	162

RESULT 14
US-08-460-458-2
: Sequence 2, Application US/08460458
: Patent No. 5738984
: GENERAL INFORMATION:
: APPLICANT: Shoscyov, Oded
: TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:

```

Query Match      27.3% Score 226; DB 1; Length 162;
Best Local Similarity 34.8%; Pred. No. 1.5e-15;
Matches 57; Conservative 31; Mismatches 58; Indels 18; Gaps 7;

QY      2  SGGVKVOYKKNDSAPGDNQIKPGIQLQVNTGSSVDLSTVTVRYWFTRDGSGSTLVYNCDW 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      3  TSSMSVEFYNSNKAQTNSITPIIKITNTSDSDLNDVKVRYVYTSDGTOGQTFW-CDH 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      62  AAMCCCG-----IRASF--GSVNPAITADTYLQLSFTGGTLAGAGSGTCE---IQNR 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62  AGALLGNSYVDNTSKVTANFVKETASP-TSPYDPYVEFGFASG--AATLKKQCFITQGR 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      109  VNKSQSNFDETDNDYSVGTNTATQDWT-KVTVYVNGRLVWGTEP 151
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119  ITRKSDMSNYQTQDNDISPDASSPTPVNPKVTGYIGGAKVLCTAP 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
US-08-460-455-2

```

: Sequence 2, Application US/08460455
: Patent No. 5837814
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: APPLICANT: Shplegli, Itai
: APPLICANT: Goldstein, Marc A.
: APPLICANT: Dol, Roy H.
: TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036

```

Query Match	27.3%	Score 226;	DB 2;	Length 162;
Best Local Similarity	34.8%	Pred. No. 1.5e-15;		
Matches	57;	Conservative	31;	Mismatches 58;
				Indels 18;
				Gaps 7;

[illegible]

Search completed: November 13, 2002, 11:56:06
Job time : 5.75718 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:55:11 : Search time 2.98707 Seconds
(without alignments)
776.473 Million cell updates/sec

Title: US-09-917-378-4

Perfect score: 829

Sequence: 1 VSGVKVQYKKNDSAPCDNQ.....TKVTYVNGRLVWGTEPSGT 154

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published_Applications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357.5	43.1	493	12	US-10-072-152-12
2	90.5	10.9	986	10	US-09-747-835A-53
3	90.5	10.9	1325	10	US-09-747-835A-24
4	90.5	10.9	1346	10	US-09-747-835A-20
5	90.5	10.9	1346	10	US-09-747-835A-61
6	90.5	10.9	1371	10	US-09-747-835A-25
7	80	9.7	1599	9	US-10-092-880-9
8	80	9.7	1600	9	US-10-092-880-10
9	75.5	9.1	1349	10	US-09-747-835A-52
10	75.5	9.1	2353	10	US-09-797-862-33
11	74.5	9.0	871	10	US-09-886-468-21
12	72.5	8.7	69	10	US-09-764-860-377
13	72.5	8.7	1333	10	US-09-815-242-10936
14	70.5	8.5	334	10	US-09-870-521-4
15	70	8.4	300	10	US-09-815-242-11359
16	70	8.4	300	10	US-09-881-752A-202
17	70	8.4	1974	9	US-09-895-913A-12
18	69	8.3	294	9	US-09-981-353-108
19	68.5	8.3	185	9	US-09-970-616-2

Query Match 43.1%; Score 357.5; DB 12; Length 493;
Best Local Similarity 42.3%; Pred. No. 4.5e-30;

ALIGNMENTS

RESULT 1

US-10-072-152-12

: Sequence 12, Application US/10072152

: Patent No. US20020142438A1

: GENERAL INFORMATION:

: APPLICANT: Andersen, Lene N.

: APPLICANT: Schulein, Martin

: APPLICANT: Lange, Niels E.

: APPLICANT: Bjornvad, Mads E.

: APPLICANT: Moller, Soren

: APPLICANT: Glad, Sanne O. S.

: APPLICANT: Kauppinen, Markus S.

: APPLICANT: Schmor, Kirk

: APPLICANT: Kongsbak, Lars

: TITLE OF INVENTION: No. US20020142438A1e1 Pectate Lyases

: FILE REFERENCE: 5378.200-US

: CURRENT APPLICATION NUMBER: US/10/072.152

: CURRENT FILING DATE: 2002-02-07

: PRIOR APPLICATION NUMBER: US/09/198.955

: PRIOR FILING DATE: 1998-11-24

: PRIOR APPLICATION NUMBER: 1343/97

: PRIOR FILING DATE: 1997-11-24

: PRIOR APPLICATION NUMBER: 1344/97

: PRIOR FILING DATE: 1997-11-24

: PRIOR APPLICATION NUMBER: 60/067,249

: PRIOR FILING DATE: 1997-12-02

: PRIOR APPLICATION NUMBER: 60/067,240

: PRIOR FILING DATE: 1997-12-02

: PRIOR APPLICATION NUMBER: 09/073,684

: PRIOR FILING DATE: 1998-05-06

: PRIOR APPLICATION NUMBER: 09/184,217

: PRIOR FILING DATE: 1998-11-02

: NUMBER OF SEQ ID NOS: 32

: SOFTWARE: FastSeq for Windows Version 4.0

: SEQ ID NO 12

: LENGTH: 493

: TYPE: PRT

: ORGANISM: Clostridium thermocellum

US-10-072-152-12


```
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 1346
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-835A-20

Query Match      10.9%  Score 90.5;  DB 10;  Length 1346;
Best Local Similarity 29.2%;  Pred. No. 0.23;
Matches 38;  Conservative 15;  Mismatches 42;  Indels 35;  Gaps 7;

Oy 31 GSSSVDLSTVTYVRYWFTRDG-----GSS--TLVYNCD-----WAAMCGGNIRASFGSVNPA 79
   | : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 414 GTPETDIDSSCSRYTLKADGTCPCSGSGTTVIYTCEFI SAYGARGSANIKVTFISVANL 473

Oy 80 TPTADTYLQLSFTGCTLAAGGSGTGEIQNRVKNK--SDWSNFDNDYSYGTNTA----FQD 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 474 TITPDPi-----SVSEQNFSIKCISDVSNYDEV-----YWNTSAGIKIYOR 515

Oy 134 WTKVTYVYVNG 143
   : | : |
Db 516 FYTTRRYLDG 525

RESULT 5
US-09-747-835A-61
; Sequence 61, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chonghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 1371
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-835A-25
```

```
; ORGANISM: Homo sapiens
US-09-747-835A-61

Query Match      10.9%  Score 90.5;  DB 10;  Length 1346;
Best Local Similarity 29.2%;  Pred. No. 0.23;
Matches 38;  Conservative 15;  Mismatches 42;  Indels 35;  Gaps 7;

Oy 31 GSSSVDLSTVTYVRYWFTRDG-----GSS--TLVYNCD-----WAAMCGGNIRASFGSVNPA 79
   | : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 414 GTPETDIDSSCSRYTLKADGTCPCSGSGTTVIYTCEFI SAYGARGSANIKVTFISVANL 473

Oy 80 TPTADTYLQLSFTGCTLAAGGSGTGEIQNRVKNK--SDWSNFDNDYSYGTNTA----FQD 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 474 TITPDPi-----SVSEQNFSIKCISDVSNYDEV-----YWNTSAGIKIYOR 515

Oy 134 WTKVTYVYVNG 143
   : | : |
Db 516 FYTTRRYLDG 525

RESULT 6
US-09-747-835A-25
; Sequence 25, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chonghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 1371
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-835A-25

Query Match      10.9%  Score 90.5;  DB 10;  Length 1371;
Best Local Similarity 29.2%;  Pred. No. 0.24;
Matches 38;  Conservative 15;  Mismatches 42;  Indels 35;  Gaps 7;

Oy 31 GSSSVDLSTVTYVRYWFTRDG-----GSS--TLVYNCD-----WAAMCGGNIRASFGSVNPA 79
   | : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 438 GTPETDIDSSCSRYTLKADGTCPCSGSGTTVIYTCEFI SAYGARGSANIKVTFISVANL 497

Oy 80 TPTADTYLQLSFTGCTLAAGGSGTGEIQNRVKNK--SDWSNFDNDYSYGTNTA----FQD 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 498 TITPDPi-----SVSEQNFSIKCISDVSNYDEV-----YWNTSAGIKIYOR 539

Oy 134 WTKVTYVYVNG 143
   : | : |
```

Db 540 FYTTRYLDG 549

RESULT 7

US-10-092-880-9

: Sequence 9, Application US/10092880

: Patent No. US20020164354A1

: GENERAL INFORMATION:

: APPLICANT: Barenkamp, Stephen J.

: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE

: FILE REFERENCE: HAEMOPHILUS

: CURRENT APPLICATION NUMBER: US/10/092,880

: CURRENT FILING DATE: 2002-03-08

: PRIOR APPLICATION NUMBER: 09/155,614

: PRIOR FILING DATE: 1998-09-30

: PRIOR APPLICATION NUMBER: 08/617,697

: PRIOR FILING DATE: 1996-04-01

: PRIOR APPLICATION NUMBER: PCT/US97/04707

: PRIOR FILING DATE: 1997-04-01

: NUMBER OF SEQ ID NOS: 11

: SOFTWARE: PatentIn ver. 2.1

: SEQ ID NO 9

: LENGTH: 1599

: TYPE: PRT

: ORGANISM: Haemophilus influenzae

US-10-092-880-9

Query Match 9.7% Score 80; DB 9; Length 1599;

Best Local Similarity 22.7%; Pred. No. 3.7;

Matches 37; Conservative 26; Mismatches 80; Indels 20; Gaps 5;

Qy 2 SGGVKVQYKKNDSAPGDNQIKPGIQLVNTGS-----SSVDLSTVTYRYWFTTRDGS-STLV 56

Db 1198 SGTVNISTKTGDIKGGIESTSCNVNITASGNTLKVSNTITGQDVT-----TADAGALTTTA 1253

Qy 57 YNCWMAAGCCGNIIRASFGSNPATPTADTYLQLSFTGCTLAAGSGTGEI-----QNRVN 110

Db 1254 GSTISATTGNANITTKTGDIKGVESSSGVTLVATGATLAVGNISGNTVTITADSGKLT 1313

Qy 111 KSDWSNFDETNDYSGYNTAFQDMTKVTYVYNGRLVWCTEPTSG 153

Db 1314 STVGSTINGTNSVTSSQSGDIEGT-----ISGNTVNVNTASTG 1351

RESULT 8

US-10-092-880-10

: Sequence 10, Application US/10092880

: Patent No. US20020164354A1

: GENERAL INFORMATION:

: APPLICANT: Barenkamp, Stephen J.

: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE

: FILE REFERENCE: HAEMOPHILUS

: CURRENT APPLICATION NUMBER: US/10/092,880

: CURRENT FILING DATE: 2002-03-08

: PRIOR APPLICATION NUMBER: 09/155,614

: PRIOR FILING DATE: 1998-09-30

: PRIOR APPLICATION NUMBER: 08/617,697

: PRIOR FILING DATE: 1996-04-01

: PRIOR APPLICATION NUMBER: PCT/US97/04707

: PRIOR FILING DATE: 1997-04-01

: NUMBER OF SEQ ID NOS: 11

: SOFTWARE: PatentIn ver. 2.1

: SEQ ID NO 10

: LENGTH: 1600

: TYPE: PRT

: ORGANISM: Haemophilus influenzae

US-10-092-880-10

Query Match 9.7% Score 80; DB 9; Length 1600;

Best Local Similarity 22.7%; Pred. No. 3.7;

Matches 37; Conservative 26; Mismatches 80; Indels 20; Gaps 5;

Qy 2 SGGVKVQYKKNDSAPGDNQIKPGIQLVNTGS-----SSVDLSTVTYRYWFTTRDGS-STLV 56

Db 1198 SGTVNISTKTGDIKGGIESTSCNVNITASGNTLKVSNTITGQDVT-----TADAGALTTTA 1253

Qy 57 YNCWMAAGCCGNIIRASFGSNPATPTADTYLQLSFTGCTLAAGSGTGEI-----QNRVN 110

Db 1254 GSTISATTGNANITTKTGDIKGVESSSGVTLVATGATLAVGNISGNTVTITADSGKLT 1313

Qy 111 KSDWSNFDETNDYSGYNTAFQDMTKVTYVYNGRLVWCTEPTSG 153

Db 1314 STVGSTINGTNSVTSSQSGDIEGT-----ISGNTVNVNTASTG 1351

RESULT 9

US-09-747-835A-52

: Sequence 52, Application US/09747835A

: Patent No. US20020146692A1

: GENERAL INFORMATION:

: APPLICANT: Yamazaki, Victoria

: APPLICANT: Tang, Y. Tom

: APPLICANT: Liu, Chenchua

: APPLICANT: Zhou, Ping

: APPLICANT: Wang, Dunrui

: APPLICANT: Zhang, Jie

: APPLICANT: Ren, Feiyan

: APPLICANT: Asundi, Vinod

: APPLICANT: Dmanac, Radoje T

: TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-L

: FILE REFERENCE: HVS-37CIP

: CURRENT APPLICATION NUMBER: US/09/747,835A

: CURRENT FILING DATE: 2002-03-08

: PRIOR APPLICATION NUMBER: US 09/729,739

: PRIOR FILING DATE: 2000-12-04

: PRIOR APPLICATION NUMBER: US 09/653,450

: PRIOR FILING DATE: 2000-08-31

: PRIOR APPLICATION NUMBER: US 09/620,312

: PRIOR FILING DATE: 2000-07-19

: PRIOR APPLICATION NUMBER: US 09/598,042

: PRIOR FILING DATE: 2000-06-20

: PRIOR APPLICATION NUMBER: US 09/552,317

: PRIOR FILING DATE: 2000-04-25

: PRIOR APPLICATION NUMBER: US 09/488,725

: PRIOR FILING DATE: 2000-01-21

: NUMBER OF SEQ ID NOS: 63

: SOFTWARE: PatentIn version 3.0

: SEQ ID NO 52

: LENGTH: 1349

: TYPE: PRT

: ORGANISM: Rattus norvegicus

US-09-747-835A-52

Query Match 9.1% Score 75.5; DB 10; Length 1349;

Best Local Similarity 30.0%; Pred. No. 8.9;

Matches 30; Conservative 10; Mismatches 37; Indels 23; Gaps 4;

Qy 31 GSSSSVDLSTVTYRYWFTTRDG-----GSS--TLVYVNC-----WAAMGCCGNIIRASFGSNPA 79

Db 412 GTPETDLESSCSTYTLKADGTQCPGSSGTTVIITCEPVSVYGARGSKNIAVTFTSVANL 471

Qy 80 TPTADTYLQLSFTGCTLAAGSGTGEIONRVNKSQNSNPFDE 119

Db 472 TITPD-----PISVSEGSQSFSTCLSDVSSPDE 499

RESULT 10

US-09-797-862-33

: Sequence 33, Application US/09797862

: Patent No. US20020102276A1

: GENERAL INFORMATION:

: APPLICANT: PEAK, IAN RICHARD ANSELM

: APPLICANT: JENNINGS, MICHAEL PAUL

APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797.862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-797-862-33

Query Match 9.18; Score 75.5; DB 10; Length 2353;
Best Local Similarity 21.98; Pred. No. 18;
Matches 44; Conservative 15; Mismatches 79; Indels 63; Gaps 6;
Qy 5 VKVOYKNNDSA-----PDQNOIKPGLQLVNTGSSSVLDLSTVTVRYWFTTRDGGSSFLVYNC 59
Db 669 LKVDNQNTDNLVTGNCNGTAVTKGFEVTKGATDADRGKVTVDATANDADKKVATVKD 728
Qy 60 DWAMCGC-----NTRASFSGVNPATPTADTYLQLSFT----- 92
Db 729 VATAINSAATEVKTENLTTSIDEDNPTDNGKDDALKAGDTLTTFKAGKNLKVKRDCKNITF 788
Qy 93 -----GCTLAAGCGSTGEIO-NRVNKSQWSNF-DETNDYSYGTNTAF 131
Db 789 DLAKNLEVKTKVSDTLTGIGTPTGGTTATPKVNITSTADGLNPAKETADAGSKN--- 845
Qy 132 QDMTKTVYVNGRLVWGTEPS 152
Db 846 -----VYLGKIATTLTPS 859

RESULT 11
US-09-886-468-21
; Sequence 21, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886.468
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061

PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-886-468-21
Query Match 9.08; Score 74.5; DB 10; Length 871;
Best Local Similarity 23.98; Pred. No. 6.4;
Matches 43; Conservative 24; Mismatches 66; Indels 47; Gaps 8;
Qy 22 KPGQLVNTGSSVDLSTVTVR---YWFTRD--GGSSTLVYNCDAAMCGC----- 67
Db 307 KTGLAISQNOEWSFTSTNTTANGGAIYATKCTLDONTTLTFTONTATATACCGGAIYVTEHD 366
Qy 68 -NIRASFSGVNPATPTADTYLQL-----SFTGGT---LAAGCGSTGEIONRVNK----- 111
Db 367 FSLKSGCTGTVTFSTNTAKTGALYSKGNSSLTGNTNLLFSGNKATGSPSSNANQEGGCA 426
Qy 112 -----SDWSNFDEN--DYSYGTNTAFDWM-----TKVTYVYVNGRLVWGTEPSCT 154
Db 427 ILAFIDSGSVSDKTLGSLTANNQEVSLTSNAATVSGGAIYATKCTLTGNSLTDFDGNATCT 486

RESULT 12
US-09-764-860-377
; Sequence 377, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764.860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 377
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-860-377

Query Match 8.78; Score 72.5; DB 10; Length 69;
Best Local Similarity 40.98; Pred. No. 0.4;
Matches 18; Conservative 3; Mismatches 12; Indels 11; Gaps 2;
Qy 30 TGSSSVLDLSTVTVRYWFTTRDGGSSSTLVYNCDAAMCGC---GNIR 70
Db 26 TPSSSSSCSNAAVRY-----TLKYHCDWFHDCGKMSGTVR 61

RESULT 13
US-09-815-242-10936
; Sequence 10936, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run On: November 13, 2002, 11:50:41 : Search time 4.71624 Seconds
(without alignments)
2058.756 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVGCRTATVYVNSDWGSGFTA.....GFNGSYSGTNTAPTTLCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	292.5	53.9	579	2 T35240	probable secreted
2	275	50.6	596	2 A55976	cellulose 1.4-beta
3	270	49.7	441	2 T12011	cellulase (EC 3.2.
4	266.5	49.1	747	2 B47093	cellulase (EC 3.2.
5	265	48.8	984	2 T44496	cellulose 1.4-beta
6	257	47.3	890	2 T35237	probable secreted
7	254	46.8	495	2 T17478	hypothetical prote
8	244.5	45.0	484	1 A24994	cellulose 1.4-beta
9	236.5	43.6	872	2 S49541	cellulase - Cellul
10	234	43.1	1045	2 A39199	endoglucanase B (E
11	230	42.4	748	2 S19652	cellodextrinase C
12	228.5	42.1	973	2 T35238	probable secreted
13	223.5	41.2	1090	2 S59077	cellulose 1.4-beta
14	213	39.2	609	2 A20073	probable chitinase
15	212.5	39.1	449	2 A24993	cellulase (EC 3.2.
16	212.5	39.1	515	2 S20493	endoglucanase - Cl
17	205	37.8	610	2 JH0573	chitinase (EC 3.2.
18	198	36.5	570	2 S56132	cellulase (EC 3.2.
19	197	36.3	400	2 T35106	probable secreted
20	197	36.3	426	2 A42360	cellulase (EC 3.2.
21	192	35.4	962	2 S03818	carboxymethylcellu
22	191.5	35.3	382	2 JC2571	cellulase (EC 3.2.
23	184.5	34.0	571	2 S13392	alpha-N-arabinofur
24	184.5	34.0	583	2 S36781	esterase D - Pseud
25	184.5	34.0	592	1 S13391	endo-1,4-beta-xyla
26	184	33.9	611	1 S06047	endo-1,4-beta-xyla
27	180	33.1	511	2 S10527	endoglucanase B pr
28	171	31.5	466	2 C42360	cellulase (EC 3.2.
29	155	28.5	142	2 A70757	hypothetical prote

30 138.5 25.5 644 1 140712 endo-1,4-beta-xyla
31 137 25.2 683 2 A82704 1,4-beta-cellobios
32 132.5 24.4 592 2 E82759 endo-1,4-beta-gluc
33 122.5 22.6 547 2 J00356 cellulase (EC 3.2.
34 121 22.3 1070 2 S75712 cellulase (EC 3.2.
35 116.5 21.5 765 2 T35719 chitinase - Strept
36 104 19.2 995 2 S50358 hypothetical prote
37 97 17.9 1778 2 T50074 probable nucleopor
38 94 17.3 335 2 T50601 endo-1,4-beta-xyla
39 93 17.1 1026 2 E86185 hypothetical prote
40 93 17.1 1026 2 A48995 paracrystalline su
41 93 17.1 1073 2 C87374 S-layer protein Rs
42 89.5 16.5 237 2 S12610 cellulase (EC 3.2.
43 88 16.2 335 2 T50600 acetylglucan estera
44 88 16.2 717 2 T06041 hypothetical prote
45 87.5 16.1 1034 2 JC2143 ice nucleation act

ALIGNMENTS

RESULT 1

T35240

probable secreted cellulase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35240

R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A:Reference number: 221572

A:Accession: T35240

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-579 <SEE>

A:Cross-references: EMBL:AL031515; PIDN:CAA20645.1; GSPDB:GN00070; SCOEIDB:SC5C7.33

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEIDB:SC5C7.33

Query Match 53.9%; Score 292.5; DB 2; Length 579;

Best Local Similarity 53.5%; Pred. No. 3e-18;

Matches 54; Conservative 14; Mismatches 26; Indels 7; Gaps 2;

Oy 4 CRATVYVNSDWGSGFTATVTNTGSRATSGTVAWSFGGNOTVTNYNTALTQSGASVT 63

Db 37 CTVDYQVNDWVGSGFTAAATVTNNGA-ATSSWSLGTAGSQKVTNSWNAKVTSQCAAVT 95

Oy 64 ATNLSYNNVIQPGOSTTFGFGNGSYSGTNTAPT-----LTC 98

Db 96 AANESYNGTSLTSGSASFGFGTGYSGSNAIPATFTLNGVTC 136

RESULT 2

A55976

cellulose 1.4-beta-cellobiosidase (EC 3.2.1.91) - Thermomonospora fusca

C:Species: Thermomonospora fusca

C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999

R:Zhang, S.; Lao, G.; Wilison, D.B.

Biochemistry 34, 3385-3395, 1995

A:Title: Characterization of a Thermomonospora fusca exocellulase.

A:Reference number: A55976; MUID:95186496; PMID:7880834

A:Accession: A55976

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-596 <2HA>

A:Cross-references: GB:U18978; NID:g664822; PIDN:AAA62211.1; PID:g664823

C:Superfamily: bacterial cellulose-binding domain homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:40-140/Domain: bacterial cellulose-binding domain homology <BCB>

F:41-139/Disulfide bonds: #status predicted

Query Match 50.6%; Score 275; DB 2; Length 596;

Db 846 ARNVGNAGVAPGASVGFGFTGSRSGTNAEP 876

RESULT 7

T17478

hypothetical protein PCZA361.11 - Amycolatopsis orientalis

C:Species: Amycolatopsis orientalis

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T17478

R:Van Wageningen, A.; Kirkpatrick, P.; Williams, D.; Harris, B.; Kershaw, J.; Lennard, N

Chem. Biol. 3, 155-162, 1998

A:Title: Sequencing and analysis of genes involved in the biosynthesis of a vancomycin g

A:Reference number: Z18804

A:Accession: T17478

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-495 <VAN>

A:Cross-references: EMBL:AJ223998; NID:e1251208; PID:e1251217; PIDN:CAA11769.1

Query Match 46.8%; Score 254; DB 2; Length 495;

Best Local Similarity 49.5%; Pred. No. 6.1e-15;

Matches 46; Conservative 16; Mismatches 31; Indels 0; Gaps 0;

Qy 2 VGRATYVYVNSDMWGSFGTATVTVTNGSRATSGWTVAMVSGNGQTVTVNYWNTALTQSGASV 61

Db 392 VDCSVAYRVDQMGNGFTASVTVTNNGTTAIDGSLRWTFENSOQVSNGNAAVRQSDQA 451

Qy 62 VTATNLSYNNVIOPCQSTTFGNGSYSGTNTAP 94

Db 452 VSAANANWTKIDPKRSVTFGFLASHSGLNPPP 484

RESULT 8

A24994

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) A precursor [invalidated] - Cellulomonas f

N:Alternate names: exo-cellobiohydrolase A

C:Species: Cellulomonas fimi

C:Date: 15-Dec-1988 #sequence_revision 22-Nov-1996 #text_change 15-Sep-2000

C:Accession: A24994

R:O'Neill, G.; Goh, S.H.; Warren, R.A.J.; Kilburn, D.G.; Miller Jr., R.C.

Gene 44, 325-330, 1986

A:Title: Structure of the gene encoding the exoglucanase of Cellulomonas fimi.

A:Reference number: A24994; MUID:87055250; PMID:3096818

A:Accession: A24994

A:Molecule type: DNA

A:Residues: 1-484 <ONE>

A:Cross-references: GB:M15824; NID:g144424; PIDN:AAA56791.1; PID:g144425

A:Note: the amino-terminal sequence of the mature protein (residues 42-71) has been dete

R:White, A.; Withers, S.G.; Gilkes, N.R.; Rose, D.R.

submitted to the Brookhaven Protein Data Bank, July 1994

A:Reference number: A67217; PDB:2EXO

A:Contents: annotation; X-ray crystallography, 1.8 angstroms, residues 42-173, 'G', 175, 'H

R:White, A.; Withers, S.G.; Gilkes, N.R.; Rose, D.R.

Biochemistry 33, 12546-12552, 1994

A:Title: Crystal structure of the catalytic domain of the beta-1,4-glycanase Cex from Ce

A:Reference number: A55905; MUID:95001978; PMID:7918478

A:Contents: annotation; X-ray crystallography, 1.8 angstroms, residues 42-173, 'G', 175, 'H

C:Genetics:

A:Gene: cex

C:Function:

A:Description: hydrolyses 1,4-beta-D-glucosidic linkages in cellulose and, releasing cel

C:Superfamily: Cellulomonas cellulose 1,4-beta-cellobiosidase A; bacterial cellulose-bi

C:Keywords: extracellular protein; glycoprotein; glycosidase; hydrolase; polysaccharide

F:1-41/Domain: signal sequence #status predicted <SIG>

F:42-484/Product: cellulose 1,4-beta-cellobiosidase #status experimental <MAT>

F:73-353/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

F:357-377/Region: 7-residue repeats (P-T-P-T-P-T-[S/T])

F:381-482/Domain: bacterial cellulose-binding domain homology <BCB>

F:387-398,403,447/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:168,274/Active site: Glu #status experimental

F:208-240,302-308/Disulfide bonds: #status experimental

F:382-481/Disulfide bonds: #status predicted

Query Match 45.0%; Score 244.5; DB 1; Length 484;

Best Local Similarity 49.5%; Pred. No. 4e-14;

Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

Qy 3 GCRATYVYVNSDMWGSFGTATVTVTNTGSHATSGWTVAMVSGNGQTVTVNYWNTALTQSGASV 62

Db 381 GCQVLGMVN-QWNTGFTANVTVKNTSSAPVDCWTLTFSPSQQVTOAWSSTVTUGSAV 439

Qy 63 TATNLSYNNVIOPCQSTTFGNGSYSGTNTAPT 95

Db 440 TVRNAPWNGSIAGGTAGQFGNGSHTGTNAAPT 472

RESULT 9

S49541

cellulase - Cellulomonas fimi

C:Species: Cellulomonas fimi

C:Date: 29-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 22-Oct-1999

C:Accession: S49541; A47093

R:Meinke, A.; Gilkes, N.R.; Kwan, E.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.

Mol. Microbiol. 12, 413-422, 1994

A:Title: Cellobiohydrolase A (CbhA) from the cellulolytic bacterium Cellulomonas fimi

A:Reference number: S49541; MUID:94344030; PMID:8065260

A:Accession: S49541

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-872 <MEI>

A:Cross-references: EMBL:L25809; NID:g456028; PIDN:AAC36898.1; PID:g456029

R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.

J. Bacteriol. 175, 1910-1918, 1993

A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cend

A:Reference number: A47093; MUID:93209933; PMID:8458833

A:Accession: A47093

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 812-872 <ME2>

A:Experimental source: ATCC 484

A:Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBI:P:128121)

C:Superfamily: bacterial cellulose-binding domain homology

F:769-870/Domain: bacterial cellulose-binding domain homology <BCB>

F:770-869/Domain: bacterial cellulose-binding domain homology <BCB>

Query Match 43.6%; Score 236.5; DB 2; Length 872;

Best Local Similarity 43.7%; Pred. No. 3.8e-13;

Matches 45; Conservative 17; Mismatches 34; Indels 7; Gaps 2;

Qy 3 GCRATYVYVNSDMWGSFGTATVTVTNTGSHATSGWTVAMVSGNGQTVTVNYWNTALTQSGASV 62

Db 769 GCTVKYSASS-WNTGFTCTVEKNGNGTAAALNGWTLGFSFADGQKYSQGSAAEWSQSQTAV 827

Qy 63 TATNLSYNNVIOPCQSTTFGNGSYSGTNTAPT-----LTCT 99

Db 828 TAKNAPWNGTLAAGSSVSIGFNGTHNGTNTAPTATPLNGVACT 870

RESULT 10

A39199

endoglucanase B (EC 3.2.1.1-) - Cellulomonas fimi

C:Species: Cellulomonas fimi

C:Date: 08-Nov-1991 #sequence_revision 08-Nov-1991 #text_change 22-Oct-1999

C:Accession: A39199

R:Meinke, A.; Braun, C.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J

J. Bacteriol. 173, 308-314, 1991

A:Title: Unusual sequence organization in CenB, an inverting endoglucanase from Cellu

A:Reference number: A39199; MUID:91100298; PMID:1987122

A:Accession: A39199

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1045 <ME1>

A:Cross-references: GB:M64644; NID:g144415; PIDN:AAA23086.1; PID:g144416

C:Superfamily: fibronectin type III repeat homology; bacterial cellulose-binding doma

C:Keywords: glycosidase; hydrolase

F:945-1045/Domain: bacterial cellulose-binding domain homology <BCB>

F:946-1044/Disulfide bonds: #status predicted

Query Match 43.1%; Score 234; DB 2; Length 1045;
Best Local Similarity 49.5%; Pred. No. 7.6e-13;
Matches 45; Conservative 14; Mismatches 30; Indels 2; Gaps 2;

OY 4 CRATYVNSDWSGSGTATVTNTGSRATSGMTVAWSFGNGQTVTNYWNTALTQSGASVT 63

DB 946 CTIVYSTNS-WNVGFTGSKVITNTGTPLT-WTLGFAFPGGQVTOGWSATNSQTCTTVT 1003

OY 64 ATNLSYNNVIOQOSTTFGFGNSYSGTNTAP 94

DB 1004 ATGLSNWATLQPGOSTDIFGNGSHPGCTNTP 1034

RESULT 11

S19652

cellodextrinase C - Pseudomonas fluorescens

C:Species: Pseudomonas fluorescens

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999

C:Accession: S19652; S16849

R: Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.

Biochem. J. 279, 793-799, 1991

A:Title: The cellodextrinase from Pseudomonas fluorescens subsp. cellulosa consists of n

A:Reference number: S19652; MUID:92061996; PMID:1953673

A:Accession: S19652

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-748 <FER>

A:Cross-references: EMBL:X61299

R: Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.

submitted to the EMBL Data Library, August 1991

A:Description: The cellodextrinase from pseudomonas fluorescens subsp cellulosa consist

A:Reference number: S16849

A:Accession: S16849

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-262, 'C', 264-291, 'K', 293-748 <FE2>

A:Cross-references: EMBL:X61299; NID:g45501; PID:g45502

C:Genetics:

A:Gene: celC

C:Superfamily: bacterial cellulose-binding domain homology; glycosidase GWGW domain hom

F:38-134/Domain: bacterial cellulose-binding domain homology <BCB>

F:183-220/Domain: glycosidase GWGW domain homology <GWG>

F:39-133/Disulfide bonds: #status predicted

Query Match 42.4%; Score 230; DB 2; Length 748;
Best Local Similarity 47.5%; Pred. No. 1.2e-12;
Matches 48; Conservative 16; Mismatches 29; Indels 8; Gaps 5;

OY 3 GCRTYVNSDWSGSGTATVTNTGSRATSGMTVAWSFGNGQTVTNYWNTALTQSGASV 62

DB 38 GCE--YVVTNSWGSFGTAAIRITNSTSSVINGMNVSQVNSNR-VTNLWNPNLGSGN-PY 93

OY 63 TATNLSYNNVIOQOSTTFGFGNSYSGTNTAPT---TCT 99

DB 94 SASLSWNGTIOPGQVTEFGFGVNSGTVNSPTVNGAACT 134

RESULT 12

T35238

probable secreted cellulase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35238

R: Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z1572

A:Accession: T35238

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-973 <SEE>

A:Cross-references: EMBL:AL031515; PIDN:CAA20643.1; GSPDB:GN00070; SCOEDB:SC5C7.31c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC5C7.31c

Query Match 42.1%; Score 228.5; DB 2; Length 973;

Best Local Similarity 48.4%; Pred. No. 2.1e-12;

Matches 44; Conservative 13; Mismatches 33; Indels 1; Gaps 1;

OY 4 CRATYVNSDWSGSGTATVTNTGSRATSGMTVAWSFGNGQTVTNYWNTALTQSGASVT 63

DB 44 CSVDYATN-DWGAGFAELTLNHRGDAIDCWTLTYAGNQKLTNMGWCTWSOSGHDVT 102

OY 64 ATNLSYNNVIOQOSTTFGFGNSYSGTNTAP 94

DB 103 VKNASYNARIACAAVSTGCGFTYSCSNAAP 133

RESULT 13

S59077

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) precursor - Cellulomonas fimi

N:Alternate names: exo-beta-1,4-cellobiohydrolase

C:Species: Cellulomonas fimi

C:Date: 15-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 22-Oct-1999

C:Accession: S59077; S59044; I40694

R: Shen, H.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J.

Biochem. J. 311, 67-74, 1995

A:Title: Cellobiohydrolase B, a second oxo-cellobiohydrolase from the cellulosytic ba

A:Reference number: S59044; MUID:96003898; PMID:7575482

A:Accession: S59077

A:Molecule type: DNA

A:Residues: 1-1090 <SHE>

A:Cross-references: EMBL:L38827; NID:g790696; PIDN:AAB00822.1; PID:g790698

A:Accession: S59044

A:Molecule type: protein

A:Residues: 54-78;456-461 <SHW>

R: Shen, H.; Tomme, P.; Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Warren, R.A.; Miller,

Biochem. Biophys. Res. Commun. 199, 1223-1228, 1994

A:Title: Stereochemical course of hydrolysis catalysed by Cellulomonas fimi CenE, a m

A:Reference number: I40694; MUID:94197708; PMID:8147863

A:Accession: I40694

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 54-106 <RES>

A:Cross-references: GB:L29042; NID:g556035; PIDN:AAA50257.1; PID:g556036

C:Genetics:

A:Gene: cbhB; cenE

C:Superfamily: bacterial cellulose-binding domain homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-53/Domain: signal sequence #status predicted <SIG>

F:54-1090/Product: cellulose 1,4-beta-cellobiosidase #status experimental <MAT>

F:989-1090/Domain: bacterial cellulose-binding domain homology <BCB>

F:990-1089/Disulfide bonds: #status predicted

Query Match 41.2%; Score 223.5; DB 2; Length 1090;
Best Local Similarity 41.0%; Pred. No. 6.6e-12;
Matches 43; Conservative 17; Mismatches 38; Indels 7; Gaps 2;

OY 1 GVCGRATYVNSDWSGSGTATVTNTGSRATSGMTVAWSFGNGQTVTNYWNTALTQSGA 60

DB 987 GGSCSVAYNASS-WNSGFTASVRITNTGTTTNGWSLGGDLTAGQKVOOGWSATWTQSGS 1045

OY 61 SVTATNLSYNNVIOQOSTTFGFGNSYSGTNTAPT-----LTCT 99

DB 1046 TVTATNAPNNGTLAPGQTVDFGNGSHTCGNPNPASFTLNGASCT 1090

RESULT 14

T42073

probable chitinase C - Streptomyces coelicolor (fragment)

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Jun-2000

C:Accession: T42073

R: Saito, A.; Fujii, T.; Yoneyama, T.; Redenbach, M.; Ohno, T.; Watanabe, T.; Miyashi

Search completed: November 13, 2002, 11:57:36
Job time : 5.71624 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:47:41 ; Search time 2.32184 Seconds
(without alignments)
1804.220 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVGCRATVYVNSDWSGFTA.....GFNGSYSGTNTAPTTLCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	399	73.5	562	1	GUN1_ACICE	P54583 acidothermu
2	270	49.7	441	1	GUN2_THERFU	P26222 thermomonas
3	266.5	49.1	747	1	GUND_CELFI	P50400 cellulomona
4	244.5	45.0	484	1	GUX_CELFI	P07986 cellulomona
5	240.5	44.3	456	1	GUNA_MICBI	P26414 microbisor
6	236.5	43.6	872	1	GUXA_CELFI	P50401 cellulomona
7	234	43.1	1045	1	GUNB_CELFI	P26225 cellulomona
8	230	42.4	748	1	GUNC_PSEFL	P27033 pseudomonas
9	225.5	41.5	880	1	GUN4_THERFU	P26221 thermomonas
10	223.5	41.2	1090	1	GUXB_CELFI	P50899 cellulomona
11	213	39.2	619	1	CHIT_STRLI	P36909 streptomyc
12	212.5	39.1	449	1	GUNA_CELFI	P07984 cellulomona
13	212.5	39.1	515	1	GUND_CLOCL	P28623 clostridium
14	205	37.8	610	1	CHIT_STRPL	P11220 streptomyc
15	202	37.2	459	1	GUNA_STRLI	P27035 streptomyc
16	192	35.4	962	1	GUNA_PSEFL	P10476 pseudomonas
17	184.5	34.0	571	1	XYNC_PSEFL	P23031 pseudomonas
18	184.5	34.0	592	1	XYNB_PSEFL	P23030 pseudomonas
19	184	33.9	611	1	XYNA_PSEFL	P14768 pseudomonas
20	180	33.1	511	1	GUNB_PSEFL	P18126 pseudomonas
21	171	31.5	466	1	GUN5_THERFU	Q01786 thermomonas
22	155	28.5	142	1	YJ87_MYCTU	Q10870 mycobacteri
23	138.5	25.5	484	1	GUNA_XANCP	P19487 xanthomonas
24	138.5	25.5	644	1	XYND_CELFI	P54865 cellulomona
25	122.5	22.6	547	1	GUN1_BUTFI	P20847 butyrivibri
26	104	19.2	995	1	Y109_YEAST	P40442 saccharomyc
27	96	17.7	1229	1	N121_HUMAN	Q9y2n3 homo sapien
28	93	17.1	1025	1	SLAP_CAUCR	P35828 caulobacter
29	91	16.8	335	1	XYNB_STRLI	P26515 streptomyc
30	89.5	16.5	237	1	GUN_ASPAC	P22669 aspergillus
31	87.5	16.1	1034	1	ICEN_PANAN	Q47879 pantoea ana
32	87.5	16.1	1258	1	ICEN_ERWHE	P16239 erwinia her
33	87.5	16.1	1322	1	ICEA_PANAN	P20469 pantoea ana

ALIGNMENTS

RESULT 1

ID	GUN1_ACICE	STANDARD;	PRT;	562 AA.
AC	P54583;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)			
DE	(Cellulase E1) (Endocellulase E1).			
OS	Acidothermus cellulolyticus.			
OC	Bacteria; Actinobacteria; Actinobacterla (class); Actinobacteridae;			
OC	Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.			
OX	NCBI_TaxID=28049;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 43068 / 11B;			
RA	Laymon R.A., Himmel M.E., Thomas S.R.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.			
RX	MEDLINE=963146058; PubMed=8718854;			
RA	Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplan P.A.;			
RT	"Crystal structure of thermostable family 5 endocellulase E1 from			
RL	Acidothermus cellulolyticus in complex with cellobiose. ";			
CC	Biochemistry 35:10648-10660(1996).			
CC	-!- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81			
CC	DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON			
CC	CARBOXYMETHYLCELULOSE.			
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	U33212; AAA75477.1; -			
PDB	1ECE; 14-OCT-96.			
DR	InterPro: IPR001919; Bac_celose-bind.			
DR	InterPro: IPR001547; GH_5.			
DR	Pfam: PF00150; cellulase; 1.			
DR	Pfam: PF00553; CBM_2; 1.			
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.			
FT	SIGNAL 1 41			
FT	CHAIN 42 562			
FT	DOMAIN 42 400			
FT	DOMAIN 401 461			
FT	DOMAIN 462 562			
FT	ACT_SITE 203 203			
FT	ACT_SITE 323 323			
FT	NUCLEOPHILE.			

```
FT DISULFID 75 161
FT DISULFID 209 212
SQ SEQUENCE 562 AA; 60747 MW; 84E6256406A35041 CRC64;

Query Match 73.5%; Score 399; DB 1; Length 562;
Best Local Similarity 69.3%; Pred. No. 2.3e-28;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;

QY 1 GVCGRATVYVNSDMSGFTATVTNTGSRATSCWTVAMSPGGNOTVTNYNTALTQSGA 60
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 462 GARTASYOVNSDMSGFTATVTNTGSRATSCWTVAMSPGGNOTVTNSNAAVTQNGQ 521
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 61 SVTATNLISYNNVIOPGOSTTGFNGSGTNTAPLTCTAS 101
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 522 SVTARNNSYNNVIOPGONTTGFQASVTGTSNAAPTACAAS 562
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I

RESULT 2
GUN2_THEFU STANDARD: PRT; 441 AA.
AC P26222;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase E-2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-2)
DE (Cellulase E-2) (Cellulase E2).
GN CELB.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=91258320; PubMed=1904434;
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RL Thermomonospora fusca.";
RN J. Bacteriol. 173:3397-3407(1991).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=YX;
RA Jung E.D., Lao G., Irwin D., Barr B., Benjamin A., Wilson D.B.;
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 32-47.
RA Wilson D.B.;
RT "Cellulases of Thermomonospora fusca.";
RL Meth. Enzymol. 160:314-323(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 32-317.
RX MEDLINE=94002001; PubMed=8399160;
RA Spezio M., Wilson D.B., Karplus P.A.;
RT "Crystal structure of the catalytic domain of a thermophilic
RL endocellulase.";
RL Biochemistry 32:9906-9916(1993).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73321; AAC06388.1; -
```

```
DR PIR; A42360; A42360.
DR PDB; 1TML; 31-JAN-94.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR001524; GH_6.
DR Pfam; PF00553; CBM_2; 1.
DR PRINTS; PR00733; GLHYDRLASE6.
DR PRODOM; PD003733; GH_6; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 31
FT CHAIN 32 441 ENDOGLUCANASE E-2.
FT DOMAIN 32 320 CATALYTIC.
FT DOMAIN 321 340 LINKER.
FT DOMAIN 341 441 CELLULOSE-BINDING.
FT ACT_SITE 110 110
FT ACT_SITE 148 148
FT ACT_SITE 296 296
FT DISULFID 111 156
FT DISULFID 263 298
FT DISULFID 346 438
FT STRAND 36 36
FT TURN 40 41
FT HELIX 43 50
FT TURN 52 53
FT TURN 55 56
FT HELIX 57 63
FT TURN 64 66
FT STRAND 69 69
FT STRAND 71 73
FT HELIX 78 95
FT TURN 96 96
FT STRAND 98 98
FT STRAND 100 103
FT TURN 109 112
FT HELIX 122 134
FT TURN 135 138
FT STRAND 142 145
FT TURN 147 148
FT HELIX 149 153
FT TURN 154 155
FT HELIX 158 178
FT TURN 180 181
FT STRAND 183 187
FT HELIX 196 205
FT TURN 206 207
FT STRAND 214 217
FT TURN 219 220
FT HELIX 225 239
FT TURN 240 240
FT TURN 242 243
FT STRAND 245 249
FT TURN 254 255
FT TURN 259 260
FT TURN 266 267
FT STRAND 275 275
FT TURN 281 282
FT STRAND 283 288
FT TURN 292 293
FT STRAND 294 294
FT TURN 302 303
FT STRAND 305 305
FT HELIX 307 315
FT TURN 316 316
SQ SEQUENCE 441 AA; 45843 MW; 87218E4537092AE5 CRC64;

Query Match 49.7%; Score 270; DB 1; Length 441;
Best Local Similarity 50.0%; Pred. No. 4.8e-17;
Matches 49; Conservative 16; Mismatches 31; Indels 2; Gaps 1;
```


RT "Insights into transition state stabilization of the beta-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants."; Nat. Struct. Biol. 5:812-818(1998).

RN [9]

RP STRUCTURE BY NMR OF 377-484.

RX MEDLINE-95284032; PubMed-7766609;

RA Xu G.-Y., Ong E., Gilkes N.R., Kilburn D.G., Muhandiram D.R., Harris-Brandts M., Carver J.P., Kay L.E., Harvey T.S.; "Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy."; Biochemistry 34:6993-7009(1995).

RN [9]

RP MUTAGENESIS OF GLU-168.

RX MEDLINE-94250681; PubMed-7910761;

RA MacLeod A.M., Lindhorst T., Withers S.G., Warren R.A.J.; "The acid/base catalyst in the exoglucanase/xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants."; Biochemistry 33:6371-6376(1994).

RL -1- FUNCTION: HYDROLYSES BOTH CELLULOSE AND XYLAN. HAS ALSO WEAK ENDOGLUCANASE ACTIVITY.

CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:

CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;

CC (2) EXOCELLULOBIODHOLASES THAT CUT THE DISSACCHARIDE CELLULOSE FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;

CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages in cellulose and cellotetraose, releasing cellobiose from the non-reducing ends of the chains.

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic linkages in xylans.

CC -1- MISCELLANEOUS: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL SITE FOR PROTEOLYSIS

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL HYDROLASES).

CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR ENBL: M15824; AA56791.1; -

DR PIR: A24994; A24994.

DR PDB: 2EXO; 08-MAR-95.

DR PDB: 1EXG; 03-JUN-95.

DR PDB: 1EXH; 03-JUN-95.

DR PDB: 1EXP; 27-JAN-97.

DR PDB: 2XYL; 18-MAR-98.

DR PDB: 2HIS; 14-OCT-98.

DR InterPro: IPR001919; Bac_cellose-bind.

DR InterPro: IPR001000; Glyco_hydro_10.

DR InterPro: IPR001230; Prenyl_site.

DR Pfam: PF00331; Glyco_hydro_10; 1.

DR Pfam: PF00553; CBM_2; 1.

DR PRINTS: PR00134; GLHYDRASE10.

DR PROSITE: PS00561; CBD_BACTERIAL; 1.

DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal;

KW Xylan degradation; Multifunctional enzyme; 3D-structure.

FT SIGNAL 1 41

FT CHAIN 42 484 EXOGLUCANASE/XYLANASE.

FT DOMAIN 42 356 CATALYTIC.

FT DOMAIN 357 376 LINKER ("HINGE") (PRO-THR BOX).

FT DOMAIN 377 484 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 168 168 PROTON DONOR.

FT ACT_SITE 274 274 NUCLEOPHILE.

FT DISULFID 208 240

FT DISULFID 302 308

FT DISULFID 382 481

FT MUTAGEN 168 168 E->A,D,G: REDUCED ACTIVITY.

SQ SEQUENCE 484 AA; 51291 MW; 6EE5486BC0E9B02F CRC64;

Query Match 45.0%; Score 244.5; DB 1; Length 484;

Best Local Similarity 49.5%; Pred. No. 9.6e-15;

Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 3 GCRTATVYVNSDWGSGFTATVTNTGSRATSGMTVANSFGNGQTNTVNTALQTQSGASV 62

DB 381 GCQVLHGVN-QWNTGFTANTYTKNTSSAPVDGWTLTTFSPSCQOVTOAMSSTVTQSGASV 439

QY 63 TATNLVSYNNVTPQSGSTTFGNGSYSGTNTAPT 95

DB 440 TVRNAPNNGSIPAGGTAQFGFNGSHTGTNAAPT 472

RESULT 5

GUNA_MICBI STANDARD; PRT; 456 AA.

AC P26414;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).

DE CELA.

GS Microbisporea bispora.

OC Bacteria; Actinobacteria; Actinobacterla (class); Actinobacteridae;

OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;

OC Thermobispora.

OX NCBI_TaxID=2006;

RN [1]

RP SEQUENCE FROM N.A.

RA Yablonsky M.D., Elliston K.O., Eveleigh D.E.;

RL (In) Coughlan M.P. (eds.);

RL Production, characterization and application of cellulose, hemicellulose and lignin enzyme degrading systems, pp.77-83, Elsevier, London (1989).

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL HYDROLASES).

DR HSSP: P26222; ITML.

DR InterPro: IPR001919; Bac_cellose-bind.

DR InterPro: IPR001524; GH_6.

DR Pfam: PF00553; CBM_2; 1.

DR Pfam: PF01341; Glyco_hydro_6; 1.

DR PRINTS: PR00733; GLHYDRASE6.

DR PRODOM: PD003733; GH_6; 1.

DR PROSITE: PS00561; CBD_BACTERIAL; 1.

DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.

DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 456 ENDOGLUCANASE A.

FT DOMAIN 31 322 CATALYTIC.

FT DOMAIN 323 355 LINKER ("HINGE") (PRO-SER BOX).

FT DOMAIN 356 456 CELLULOSE-BINDING.

FT ACT_SITE 113 113 BY SIMILARITY.

FT ACT_SITE 151 151 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 300 300 NUCLEOPHILE (BY SIMILARITY).

FT DISULFID 114 159 BY SIMILARITY.

FT DISULFID 267 302 BY SIMILARITY.

FT DISULFID 360 453 BY SIMILARITY.

SQ SEQUENCE 456 AA; 47011 MW; B06D8595E322848F CRC64;

Query Match 44.3%; Score 240.5; DB 1; Length 456;

DR	Pfam:	PF00553;	CBM_2; 1.
DR	Pfam:	PF01341;	Glyco_hydro_6; 1.
DR	PRINTS:	PR0014;	FNTYPEI11.
DR	PRINTS:	PR00733;	GLHYDRLASE6.
DR	ProDom:	PD003733;	GH_5; 2.
DR	SMART:	SM00060;	FN3; 3.
DR	PROSITE:	PS00561;	CBD_BACTERIAL; 1.
DR	PROSITE:	PS00655;	GLYCOSYL_HYDROL_F6_1; 1.
DR	PROSITE:	PS00656;	GLYCOSYL_HYDROL_F6_2; 1.
KW	Cellulose degradation: Hydrolase; Glycosidase; Repeat; Signal.		
FT	SIGNAL	1..40	
FT	CHAIN	41..872	EXOGLUCANASE A.
FT	DOMAIN	41..477	CATALYTIC.
FT	DOMAIN	478..563	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	573..664	FIBRONECTIN TYPE-III 2.
FT	DOMAIN	673..768	FIBRONECTIN TYPE-III 3.
FT	DOMAIN	769..872	CELLULOSE-BINDING (BY SIMILARITY).
FT	ACT_SITE	188..188	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	410..410	NUCLEOPHILE (BY SIMILARITY).
FT	DISULFID	140..202	BY SIMILARITY.
FT	DISULFID	374..428	BY SIMILARITY.
FT	DISULFID	770..869	BY SIMILARITY.
SQ	SEQUENCE	872 AA; 89300 MW; 7883B407F995533B CRC64;	
	Query Match	43.6%; Score 236.5; DB 1; Length 872;	
	Best Local Similarity	43.7%; Pred. No. 9e-14;	
	Matches	45; Conservative 17; Mismatches 34; Indels 7; Gaps 2;	
QY	3	GCRTATVNVSDMGSGGTATVTGTTNGSRATSGMTVAWSFGCGNQTVMYWNLTALTOCSAV	62
Db	769	CTIVKYSASS-WNTGPTGTVEVKNNGTAALNGWTLGFSFADGQKYSCNSQSQTAV	827
QY	63	TATNLSYNNVIQPGQSTTFPGFNGSYSCTNTAPT-----LTCT	99
Db	828	TAKNPAPNGTFLAAGSSVSIGFNGTHNGTNTAPTFTLNGVACT	870
RESULT 7			
GUNB_CELFI			
ID	GUNB_CELFI	STANDARD;	PRT: 1045 AA.
AC	P26225;		
DT	01-MAY-1992 (Rel. 22, Created)		
DT	01-MAY-1992 (Rel. 22, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase B)		
GN	(Cellulase B).		
OS	Cellulomonas fimi.		
OC	Bacteria; Actinobacteria; Actinobacterla (class); Actinobacteridae;		
OC	Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.		
OX	NCBI_TaxId=1708;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=g1100298; PubMed=19871122;		
RA	Meinke A., Braun C., Gilkes N.R., Kilburn D.G., Miller R.C. Jr.,		
RA	Warren R.A.J.;		
RT	"Unusual sequence organization in CenB, an inverting endoglucanase		
RT	from Cellulomonas fimi.";		
RL	J. Bacteriol. 173:308-314(1991).		
RN	[2]		
RP	DOMAINS.		
RX	MEDLINE=g2041609; PubMed=1938913;		
RA	Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;		
RT	"Multiple domains in endoglucanase B (CenB) from Cellulomonas fimi:		
RT	functions and relatedness to domains in other polypeptides.";		
CC	-1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE		
CC	GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:		
CC	(1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;		
CC	(2) EXOCELLOBIOTRYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE		
CC	FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;		
CC	(3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER		

```

CC CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- MISCELLANEOUS: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A
CC POTENTIAL SITE FOR PROTEOLYSIS.
CC -!- MISCELLANEOUS: MAY CONTAIN A SECOND CBD IN THE CATALYTIC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: M64644; AAA23086.1; -.
CC PIR: A39199; A39199.
CC HSSP: P26221; ITR4.
CC InterPro: IPR001919; Bac_cellose-bind.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR001701; CH_9.
CC Pfam: PF00441; fn3_3.
CC Pfam: PF00553; CBM_2; 1.
CC Pfam: PF00759; Glyco_hydro_9; 1.
CC Pfam: PF00942; CBM_3; 1.
CC SMART: SM00060; FN3_3.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
KW SIGNAL 1 33
FT CHAIN 34 1045
FT DOMAIN 34 643
FT DOMAIN 644 650
FT DOMAIN 651 733
FT DOMAIN 734 748
FT DOMAIN 749 830
FT DOMAIN 831 846
FT DOMAIN 847 930
FT DOMAIN 931 944
FT DOMAIN 945 1045
FT ACT_SITE 410 410
FT ACT_SITE 449 449
FT ACT_SITE 458 458
FT DISULFID 946 1044
FT SEQUENCE 1045 AA; 108990 MW; AC2P7B84E4E3C4F0 CRC64;

Query Match 43.1%; Score 234; DB 1; Length 1045;
Best Local Similarity 49.5%; Pred. No. 1.8e-13;
Matches 45; Conservative 14; Mismatches 30; Indels 2; Gaps 2;

QY 4 CRATYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGNGQTVTNWNTALTQSGASVT 63
DB 946 CTVVSTNS-WNVGFTGSKVINTGTPTLT-WTLGFAFPGGOQVWGWSATWSQTGTTVT 1003

QY 64 ATNLNNVNIQPCQSTTFGNGSYGCTNAP 94
DB 1004 ATGLSNWNLQPCQSTDIGFNGCHPCTNTP 1034

RESULT 8
GUNC_PSEFL
ID GUNC_PSEFL STANDARD; PRT; 748 AA.
AC P27033;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Endoglucanase C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (EGC) (Cellodextrinase C).
GN CELC.

```

```

OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-47.
RC STRAIN=Sp. Cellulosa;
RA MEDLINE=92061996; PubMed=1953673;
RA Ferreira L.M.A., Hazlewood G.P., Barker P.J., Gilbert H.J.;
RT "The cellodextrinase from Pseudomonas fluorescens subsp. cellulosa
RT consists of multiple functional domains.";
RL Biochem. J. 279:793-799(1991).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD)
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: X61299; CAA43597.1; -.
CC PIR: S16849; S16849.
CC PIR: S19652; S19652.
CC HSSP: P07986; IEXG.
CC InterPro: IPR001919; Bac_cellose-bind.
CC InterPro: IPR002883; CBD_5.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF00553; CBM_2; 1.
CC Pfam: PF02013; CBM_10; 1.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW SIGNAL 1 38
FT CHAIN 39 748
FT DOMAIN 39 136
FT DOMAIN 137 179
FT DOMAIN 228 280
FT DOMAIN 281 748
FT DISULFID 39 133
FT ACT_SITE 503 503
FT ACT_SITE 653 653
FT SEQUENCE 748 AA; 80197 MW; AC96104137932B76 CRC64;

Query Match 42.4%; Score 230; DB 1; Length 748;
Best Local Similarity 47.5%; Pred. No. 2.9e-13;
Matches 48; Conservative 16; Mismatches 29; Indels 8; Gaps 5;

QY 3 GCRATYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGNGQTVTNWNTALTQSGASV 62
DB 38 GCE--VYVVTNSMGSGPTAAIRITNSTSSVINGWNVSWQVNSNR-VTNLWNPNLGSGN-PY 93

QY 63 TATNLNNVNIQPCQSTTFGNG-SYSGTNTAPT-----TCT 99
DB 94 SASNLWSNGTIQPGQTVTFGFGVTSPTVNGAACT 134

RESULT 9
GUNA_THEFU
ID GUNA_THEFU STANDARD; PRT; 880 AA.
AC P26221; Q08167;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase E-4 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-4)
DE (Cellulase E-4) (Cellulase E4).

```

GN CELD.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=94028932; PubMed=8215374;
RA Jung E.D., Lao G., Irwin D., Barr B.K., Benjamin A., Wilson D.B.;
RT "DNA sequences and expression in Streptomyces lividans of an
exoglucanase gene and an endoglucanase gene from Thermomonospora
fusca.";
RL Appl. Environ. Microbiol. 59:3032-3043(1993).
RN [2]
RP REVISIONS.
RA Wilson D.B.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=91258320; PubMed=1904434;
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
Thermomonospora fusca.";
RL J. Bacteriol. 173:3397-3407(1991).
RN [4]
RP SEQUENCE OF 47-67.
RA Wilson D.B.;
RT "Cellulases of Thermomonospora fusca.";
RL Meth. Enzymol. 160:314-323(1988).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 47-651.
RX MEDLINE=97475222; PubMed=9334746;
RA Sakon J., Irwin D., Wilson D.B., Karplus P.A.;
RT "Structure and mechanism of endo/exocellulase E4 from Thermomonospora
fusca.";
RL Nat. Struct. Biol. 4:810-818(1997).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L20093; AAB42155.1;
DR EMBL: M73322; AAA27397.1; ALT_SEQ.
DR IIR; B42360; B42360.
DR IIR; B42360; B42360.
DR PDB: 1J54; 17-SEP-97.
DR PDB: 1TF4; 04-SEP-97.
DR PDB: 3TF4; 04-SEP-97.
DR PDB: 4TF4; 04-SEP-97.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FN_III_repeat.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR001230; Prenyl_site.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR Pfam: PF00942; CBM_3; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.

DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 46
FT CHAIN 47 880 ENDOGLUCANASE E-4.
FT DOMAIN 776 880 CELLULOSE-BINDING.
FT ACT_SITE 427 427 BY SIMILARITY.
FT ACT_SITE 461 461 BY SIMILARITY.
FT ACT_SITE 470 470 BY SIMILARITY.
SQ SEQUENCE 880 AA: 95202 MW: 5EA9A6ABF45A4D9A CRC64;
Query Match 41.5%; Score 225.5; DB 1; Length 880;
Best Local Similarity 45.7%; Pred. No. 8.6e-13;
Matches 42; Conservative 17; Mismatches 32; Indels 1; Gaps 1;
QY 4 CRATVYVNDWGSFGTATVTNTGSRATSGTWTGVSFGGNTVTNYNNTALTQSGASYT 63
DB 778 CTVGYSTN-DWDSGFTASIRITYHCTAPLSSWELSFTEPAGQOVTHGNATWRDCAAVT 836
QY 64 ATNLSYNNVIOPGOSTTTFGNGSYSGTNTAPT 95
DB 837 ATPMSWSSLAPGATVEVGFNGSWGSGNTPTPT 868
RESULT 10
GUXB_CELFI STANDARD: PRT; 1090 AA.
ID GUXB_CELFI
AC PS0899;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
DE (1,4-beta-cellobiohydrolase B) (CBP120).
GN CBHB OR CENE.
OS Cellulomonas flm1.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC STRAIN=ATCC 484;
RX MEDLINE=96003898; PubMed=7575482;
RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulobiohydrolase B, a second exo-cellobiohydrolase from the
cellulolytic bacterium Cellulomonas flm1.";
RL Biochem. J. 311:67-74(1995).
RN [2]
RP SEQUENCE OF 54-75.
RX MEDLINE=93209933; PubMed=8458833;
RA Melnik A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase
D (Cend), a family A beta-1,4-glucanase.";
RL J. Bacteriol. 175:1910-1918(1993).
RN [3]
RP SEQUENCE OF 54-78.
RX MEDLINE=94197708; PubMed=8147863;
RA Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
RA Warren R.A.J., Miller R.C. Jr.;
RT "Stereochemical course of hydrolysis catalysed by Cellulomonas flm1
CenE, a member of a new family of beta-1,4-glucanases.";
RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC -1- FUNCTION: HYDROLYSE CELLOHEXAOSE TO A MIXTURE OF CELLOTETRAOSE,
CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT
HYDROLYSED CELLOPENTAOSE TO CELLOTRIOSE AND CELLOBIOSE, AND
CELLOTETRAOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTRIOSE.
CC HAS ALSO WEAK ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS
WITH INVERSION OF ANOMERIC CONFIGURATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
in cellulose and cellotetraose, releasing cellobiose from the non-
reducing ends of the chains.
CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
(CBD).

```
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L38827; AAC00822.1; -.
CC HSSP: P07986; 1EXG.
CC InterPro: IPR001919; Bac_cellose-bind.
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR003962; FN.III-repeat.
CC InterPro: IPR000556; Glyco_hydro_48.
CC Pfam: PF00041; fn3; 3.
CC Pfam: PF00553; CBM_2; 1.
CC PRINTS: PR0011; Glyco_hydro_48; 1.
CC PRINTS: PR0014; FNTYPEIIL.
CC PRINTS: PR00844; GLYVDRLASE48.
CC ProDom: PD011903; Glyco_hydro_48; 1.
CC SMART: SM00060; FN3; 3.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC Cellulose degradation: Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 33
FT PROPEP 34 53
FT CHAIN 54 1090
FT DOMAIN 54 699
FT DOMAIN 700 785
FT DOMAIN 794 884
FT DOMAIN 891 978
FT DOMAIN 989 1090
FT ACT_SITE 513 513
FT DISULFID 990 1089
FT SEQUENCE 1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;
CC -----
Query Match 41.2%; Score 223.5; DB 1; Length 1090;
Best Local Similarity 41.0%; Pred. No. 1.6e-12;
Matches 43; Conservative 17; Mismatches 38; Indels 7; Caps 2;
CC 1 GVCGRATYVNVDSGSGFTATVTTNGSRATSGWTVAMSGNGQVNTVNTALQTQSGA 60
CC 987 GCGSAYNASS-WNSGFTASVITNTGTTINGSLGDLTAGKQVQGSATWTQSGS 1045
CC 61 SVTATLSYNNVIOQSGSTFFGNGSYSGTNTAPT-----LTCT 99
CC 1046 TVTATNAPNNGTLAPQCTVDVNGSHTGQNPASPFTLNGASCT 1090
CC -----
RESULT 11
CHIT_STRLI STANDARD; PRT; 619 AA.
AC P36909;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Chitinase C precursor (EC 3.2.1.14).
GN CHIC.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93294525; PubMed=8515228;
RA Fujii T., Miyashita K.;
RT "Multiple domain structure in a chitinase gene (chic) of Streptomyces
RL J. Gen. Microbiol. 139:677-686(1993).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
```

```
CC -!- INDUCTION: BY CHITIN.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D12647; BAA02168.1; -.
CC HSSP: P07986; 1EXG.
CC InterPro: IPR001919; Bac_cellose-bind.
CC InterPro: IPR001579; Chitinase_18/2.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR001223; Glyco_hydro_18.
CC Pfam: PF00041; fn3; 1.
CC Pfam: PF00553; CBM_2; 1.
CC ProDom: PD000471; Glyco_hydro_18; 1.
CC SMART: SM00060; FN3; 1.
CC PROSITE: PS00018; EF_HAND; UNKNOWN_1.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC PROSITE: PS01095; CHITINASE_18; 1.
CC Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal.
FT SIGNAL 1 30
FT CHAIN 31 619
FT DOMAIN 35 140
FT DOMAIN 148 230
FT DOMAIN 240 619
FT ACT_SITE 382 382
FT SEQUENCE 619 AA; 65200 MW; A23CEE5B3C5D6F21 CRC64;
CC -----
Query Match 39.2%; Score 213; DB 1; Length 619;
Best Local Similarity 45.9%; Pred. No. 7.6e-12;
Matches 39; Conservative 13; Mismatches 33; Indels 0; Caps 0;
CC 6 ATTVVNSDMSGSGFTATVTTNGSRATSGWTVAMSGNGQVNTVNTALQTQSGAVTAT 65
CC 36 ATAKYSDMGTGCGSWTVANTGTTSLSSWTVEVDFTGTCTKVTSAWDATVNTSGDHTAK 95
CC 66 NLSYNNVIOQSGSTFFGNGSYSGT 90
CC 96 NVGWNGLAPGASVSGFGNGSGPCS 120
CC -----
RESULT 12
GUNA_CELFI STANDARD; PRT; 449 AA.
AC P07984;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
GN CELULASE.
OS Cellulomonas fimi.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Micrococineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87055249; PubMed=3023193;
RA Wong W.K.R., Gerhard B., Guo Z.M., Kilburn D.G., Warren R.A.J.,
RA Miller R.C. Jr.;
RT "Characterization and structure of an endoglucanase gene cenA of
RT Cellulomonas fimi.";
```


Gene 44:315-324 (1986).
[2]
RN DOMAINS.
RP MEDLINE=90036847; PubMed=2681184;
RA Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RX "Structural and functional analysis of a bacterial cellulase by
RT proteolysis."; J. Biol. Chem. 264:17802-17808(1989).
RL [3]
RN DISULFIDE BONDS.
RP MEDLINE=92104156; PubMed=1761039;
RA Gilkes N.R., Clayssens M., Aebersold R., Henrissat B., Meinke A.,
RX Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;
RA "Structural and functional relationships in two families of beta-1,4-
RT glycanases."; Eur. J. Biochem. 202:367-377(1991).
RL [3]
RN FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC -1- GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBIODOLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -2- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -3- PWM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -4- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -5- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M15823; AAA23084.1; -.
DR PIR: A24993; A24993.
DR HSP: P07986; IEXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001524; GH_6.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLYHYDRLASE6.
DR PRODOM: PD003733; GH_6; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR Cellulose degradation: Hydrolase: Glycosidase; Signal.
KW SIGNAL 1 31
FT CHAIN 32 449 ENDOGLUCANASE A.
FT DOMAIN 32 138 CELLULOSE-BINDING.
FT DOMAIN 139 168 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 438 449 CATALYTIC.
FT ACT_SITE 247 247 BY SIMILARITY.
FT ACT_SITE 283 283 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 423 423 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 35 134
FT DISULFID 248 291
FT DISULFID 390 426
SQ SEQUENCE 449 AA; 67FF887814B3348D CRC64;

Query Match 39.1%; Score 212.5; DB 1; Length 449;
Best Local Similarity 37.1%; Pred. No. 6.1e-12;
Matches 39; Conservative 17; Mismatches 42; Indels 7; Gaps 2;

QY 3 GCRATYVNSDMGSGFATVTVTNTGSRATSGHTVAHSGFGNQTVTYNWTALTQSGASV 62
Db 34 GCRVDYAVTNQWPGFGFANYTINLGD -PVSSMKLDWYTAGRIQIQLMNGTASTNGQV 92

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 10:50:41 : Search time 9.35991 Seconds
(without alignments)
2223.392 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVCGRATVYVNSDMSGFTA.....GFNGSYSGTNTAPTLTCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirois:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	292.5	53.9	579	16	O86730 streptomyc
2	275	50.6	596	2	Q60029 thermomonos
3	265	48.8	984	2	Q9XCD4
4	257	47.3	890	16	O86727 streptomyc
5	254	46.8	495	2	O52799 amycolatops
6	247.5	45.6	881	2	Q9AF09 pseudomonas
7	244.5	45.0	485	2	O59277 cellulomona
8	235	43.3	623	2	O8ROP6 streptomyc
9	228.5	42.1	381	2	Q54331 streptomyc
10	228.5	42.1	381	16	Q9RJY3 streptomyc
11	228.5	42.1	973	16	O86728 streptomyc
12	224	41.3	629	2	Q9L8C0 streptomyc
13	220	40.5	974	2	O08166 thermomonos
14	213	39.2	270	2	Q9S675 streptomyc
15	213	39.2	609	16	Q9Z9M8 streptomyc
16	204.5	37.7	377	2	O08468 streptomyc

17	198	36.5	570	2	Q59665	Q59665 pseudomonas
18	197	36.3	400	16	O69962	O69962 streptomyc
19	197	36.3	649	2	Q9F7L3	Q9F7L3 pseudomonas
20	195.5	36.0	450	16	Q9RJ91	Q9RJ91 streptomyc
21	192.5	35.5	482	2	P74912	P74912 thermobifid
22	191.5	35.3	382	2	O59963	O59963 streptomyc
23	189	34.8	364	16	Q9RJY2	Q9RJY2 streptomyc
24	187.5	34.5	364	2	O07862	O07862 streptomyc
25	187	34.4	486	16	Q9F3Q0	Q9F3Q0 streptomyc
26	186	34.3	358	2	Q9X5K8	Q9X5K8 streptomyc
27	184.5	34.0	583	2	O51815	O51815 pseudomonas
28	178.5	32.9	457	2	O53488	O53488 micromonos
29	162.5	29.9	371	2	Q9KIH1	Q9KIH1 streptomyc
30	161.5	29.7	461	2	O59922	O59922 streptomyc
31	156	28.7	571	2	O9EY44	O9EY44 cellulomona
32	144	26.5	745	2	Q9K5C7	Q9K5C7 clavibacter
33	137	25.2	683	16	Q9PDW2	Q9PDW2 xyliella fas
34	134.5	24.8	335	2	Q9RQB8	Q9RQB8 cellulomona
35	133.5	24.6	635	2	O66065	O66065 fibrobacter
36	132.5	24.4	592	16	Q9PF60	Q9PF60 xyliella tas
37	131	24.1	674	2	Q9FAC8	Q9FAC8 bacillus ce
38	129	23.8	635	2	O07088	O07088 bacillus th
39	125	23.0	481	2	O86064	O86064 actinomycs
40	125	23.0	727	2	Q9AF65	Q9AF65 clavibacter
41	123.5	22.7	384	2	Q9X602	Q9X602 streptomyc
42	123	22.7	676	2	Q93AD3	Q93AD3 bacillus th
43	121	22.3	1070	16	O55365	O55365 synethocyst
44	118.5	21.8	489	16	Q9F312	Q9F312 streptomyc
45	116.5	21.5	765	16	O54183	O54183 streptomyc

ALIGNMENTS

RESULT 1

O86730	PRELIMINARY:	PRT:	579 AA.
ID O86730			
AC O86730:			
DT 01-NOV-1998 (TREMBlrel. 08, Created)			
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)			
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)			
DE Putative secreted cellulase.			
GN SC06548 OR SC5C7.33.			
OS Streptomyces coelicolor.			
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycs.			
OX NCBI_TaxID=1902;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=A3(2);			
RA Seeger K.J., Harris D.;			
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=A3(2);			
RA Parkhill J., Barrell B.G., Rajandream M.A.;			
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
RN [3]			
RP SEQUENCE FROM N.A.			
RC STRAIN=A3(2);			
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,			
RT "A set of ordered cosmids and a detailed genetic and physical map for			
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";			
RL Mol. Microbiol. 21:77-96(1996).			
RN [4]			
RP SEQUENCE FROM N.A.			
RC STRAIN=A3(2) / M145;			
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,			
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,			
Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,			
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,			


```
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE-97000351; PubMed-8843436;
RA Redenbach M., Kleser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warten T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL031515; CAA20642.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR002860; GH_BNR.
DR Pfam: PF02012; BNR; 10.
DR Pfam: PF00553; CBM_2; 1.
SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;

Query Match 47.3%; Score 257; DB 16; Length 890;
Best Local Similarity 51.6%; Pred. No. 4; 1e-13;
Matches 47; Conservative 10; Mismatches 34; Indels 0; Gaps 0;

QY 4 CRYATVYVNSDWSGFTATVTVTNGSRATSGWTVAMSGFNGQTVTVYNTALTOSGASVT 63
Db 786 CEVTVTVTNPQPGFQADVRLTNGTSAWNGNSLDWFSFGQGVTRMNAERTQAGT/5VT 845
QY 64 ATNLSYNNVIOQCSTTFGNGSYSGTNTAP 94
Db 846 ARNVGNAGVAPGASVGFGTGSRSGTNAEP 876

RESULT 5
O52799 PRELIMINARY; PRT; 495 AA.
AC O52799;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PCZA361.11.
OS Anycolatopsis orientalis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae; Anycolatopsis.
OX NCBI_TaxID=31958;
RN [1]
RP SEQUENCE FROM N.A.
RA van Wageningen A., Kirkpatrick P., Williams D., Harris B., Kershaw J.,
RA Lennard N., Jones M., Jones S., Solenberg P.;
RT "Sequencing and analysis of genes involved in the biosynthesis of a
RT vancomycin group antibiotic.";
RL Chem. Biol. 3:155-162(1998).
DR EMBL: AJ23998; CAA11769.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR Pfam: PF00553; CBM_2; 1.
SQ SEQUENCE 495 AA; 52516 MW; 8F54AF6C4761662A CRC64;
```

```
Query Match 46.8%; Score 254; DB 2; Length 495;
Best Local Similarity 49.5%; Pred. No. 3; 9e-13;
Matches 46; Conservative 16; Mismatches 31; Indels 0; Gaps 0;

QY 2 VGRATVYVNSDWSGFTATVTVTNGSRATSGWTVAMSGFNGQTVTVYNTALTOSGAS 61
Db 392 VDCSVAYKVDQWQNGFTASVTVTNGCTTAIDGSLRWTFENSQOVSNAGNAAVRQSDQA 451
QY 62 VTATNLSYNNVIOQCSTTFGNGSYSGTNTAP 94
Db 452 VSAANWNTKIDPKSVTFGLASHSLNPPP 484

RESULT 6
O9AF09 PRELIMINARY; PRT; 881 AA.
AC O9AF09;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Rhamnogalacturonan lyase.
OS Pseudomonas cellulosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=155077;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A1CC 55703;
RX MEDLINE-21157041; PubMed=11256961;
RA McKie V.A., Vincken J.P., Voragen A.G.J., Van Den Broek L.A.M.,
RA Stimson E., Gilbert H.J.;
RT "A new family of rhamnogalacturonan lyases contains an enzyme that
RT binds to cellulose.";
RL Biochem. J. 355:167-177(2001).
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL: AY026755; AAK20911.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR003961; FN_III.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00041; fn3; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
KW Hydrolase; Lyase.
SQ SEQUENCE 881 AA; 93963 MW; E8792769751FEEC7 CRC64;

Query Match 45.6%; Score 247.5; DB 2; Length 881;
Best Local Similarity 45.7%; Pred. No. 2; 4e-12;
Matches 42; Conservative 21; Mismatches 26; Indels 3; Gaps 2;

QY 7 TYVNSDWSGFTATVTVTNGSRATSGWTVAMSGFNGQTVTVYNTALTOSGAS-VTAT 65
Db 787 SYVITNMGSGFTCAIRITNRGSSAINGVNVSWTYSGNTRISNSWNA--TVSGSNPYSAA 844
QY 66 NLSYNNVIOQCSTTFGNGSYSGTNTAPTLT 97
Db 845 NLGNATIQGVTFQFGQGTSGTETPVIS 876

RESULT 7
O59277 PRELIMINARY; PRT; 485 AA.
AC O59277;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Exo-beta-1,4-glucanase precursor (EC 3.2.1.91).
GN CEX.
OS Cellulomonas fimi.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
```

```
RN RP SEQUENCE FROM N.A.
RX MEDLINE=87055250; PubMed=3096818;
RA O'Neill G., Goh S.H., Warren R.A., Kilburn D.G., Miller R.C. Jr.;
RT "Structure of the gene encoding the exoglucanase of Cellulomonas
RL fimi.";
RL Gene 44:325-330(1986).
RN RP SEQUENCE FROM N.A.
RA Ramirez C., MacLeod A.M., Gillies N.R.;
RT "Revision to the Cellulomonas fimi Cex gene sequence encoding an exo-
RT beta-1,4-glucanase and beta-1,4-xylanase.";
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L11080; AAA56792.1; -.
DR HSP; P07986; 1EXP.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR001230; Prenyl_site.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS00591; GLYCOSYL-HYDROL-F10; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 42 POTENTIAL.
FT CHAIN 43 485 BETA-1,4-XYLANASE.
SQ SEQUENCE 485 AA; 51209 MW; 1E0BFD54C0B95A27 CRC64;

Query Match 45.0%; Score 244.5; DB 2; Length 485;
Best Local Similarity 49.5%; Pred. No. 2.3e-12;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 3 GCRATVYVNSDMSGFTATVTNTGSRATSGWTVAWSFGGNOTVNYNNTALTQSGASV 62
DB 382 GCOVLGKV-NQNTGFANVTVKNTSSAPVDGWTLTFFSPSGQOVTQANSSVTQSGSAV 440

QY 63 TATNLSYNNVIOPGQSTTFGFGNSYSGTNTAPT 95
DB 441 TVRNAPWNGSIPAGCTAOFPGFNGSHGTGNTAAPT 473

RESULT 8
Q8RQP6 PRELIMINARY: PRT; 623 AA.
AC Q8RQP6;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE Chitinase III (EC 3.2.1.14) (Fragment).
GN CHI 111.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HUT 6037;
RA Watanabe T.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HUT 6037;
RA Kawase T., Kanai R., Ohno T., Tanabe T., Nikaido N., Miyashita K.,
RA Mitutomi M., Watanabe T.;
RT "Identification of three family 18 chitinase genes of Streptomyces
RT griseus HUT6037.";
RL Chitin and chitosan research 7:241-251(2001).
DR EMBL; AB081807; BAB86377.1; -.
KW Hydrolase; Glycosidase.
FT NON_TER 1
FT CHAIN 623 623
SQ SEQUENCE 623 AA; 64893 MW; 02AAAB520A6C7E1C CRC64;
```

```
Query Match 43.3%; Score 235; DB 2; Length 623;
Best Local Similarity 50.6%; Pred. No. 1.8e-11;
Matches 44; Conservative 9; Mismatches 34; Indels 0; Gaps 0;

QY 6 ATYVNSDMSGFTATVTNTGSRATSGWTVAWSFGGNOTVNYNNTALTQSGASVAT 65
DB 53 ATYTKSDMSGFGQWTKNTCTTALSSWTIEMDPPSGTTAGSAMDATLTSSGGHYTAK 112

QY 66 NLSYNNVIOPGQSTTFGFGNSYSGTNT 92
DB 113 NLSWNGSVAPGASISFGFNGTSGSPT 139

RESULT 9
Q54331 PRELIMINARY: PRT; 381 AA.
ID Q54331;
AC Q54331;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Cellulase B precursor.
GN CELB.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=66;
RX MEDLINE=94288649; PubMed=8017952;
RA Witmann S., Sharek F., Kluepfel D., Morosoli R.;
RT "Purification and characterization of the CelB endoglucanase from
RT Streptomyces lividans 66 and DNA sequence of the encoding gene.";
RL Appl. Environ. Microbiol. 60:1701-1703(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=66;
RA Sharek F.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U04629; AAB71950.1; -.
DR HSP; P07986; 1EXP.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR002594; GH_12.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; GH_12; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 381 CELLULASE B.
SQ SEQUENCE 381 AA; 39239 MW; A7E99BF590FA24EC CRC64;

Query Match 42.1%; Score 228.5; DB 2; Length 381;
Best Local Similarity 48.9%; Pred. No. 3.6e-11;
Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;

QY 4 CRATVYVNSDMSGFTATVTNTGSRATSGWTVAWSFGGNOTVNYNNTALTQSGASVT 63
DB 279 CAVSYGTNV-WQDGFADVTVTNTGTPVDGWOLATLPSGQRITNANASLTSPSGSVT 337

QY 64 ATNLSYNNVIOPGQSTTFGFGNSYSGTNTAPT 95
DB 338 ATGASHNARIAPGSLSFQFGCTGYGCAFAEPT 369

RESULT 10
Q9RJY3 PRELIMINARY: PRT; 381 AA.
ID Q9RJY3;
AC Q9RJY3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
```


DE	Putative secreted cellulase B.
GN	CELB OR SC01187 OR SCGL1A.18.
OS	Streptomyces coelicolor.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX	NCBI_TaxID=1902;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	SEQUENCE-A3(2);
RA	Oliver K., Harris D.;
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[2]	
RP	SEQUENCE FROM N.A.
RC	STRAIN-A3(2);
RA	Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[3]	
RP	SEQUENCE FROM N.A.
RC	STRAIN-A3(2);
RX	MEDLINE=97000351; PubMed=8843436;
RA	Redenbach M., Kieser H.M., Denapajta D., Eichner A., Cullum J.,
RA	Kinashi H., Hopwood D.A.;
RT	"A set of ordered cosmids and a detailed genetic and physical map for
RT	the 8 mb Streptomyces coelicolor A3(2) chromosome.";
RL	Mol. Microbiol. 21:77-96(1996).
[4]	
RP	SEQUENCE FROM N.A.
RC	STRAIN-A3(2) / M45;
RA	Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA	Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA	Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA	Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA	Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA	Hopwood D.A.;
RT	"Complete genome sequence of the model actinomycete Streptomyces
RT	coelicolor A3(2).";
RT	Nature 417:141-147(2002).
DR	EMBL: ALJ3210 - CAB61599.1; -
DR	HSP: P07986; IEXG.
DR	InterPro: IPR001919; Bac_cellose-bind.
DR	InterPro: IPR002594; GH_12.
DR	Pfam: PF00553; CBM_2; 1.
DR	Pfam: PF01670; Glyco_hydro_12; 1.
DR	ProDom: PD004316; GH_12; 1.
DR	PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
SO	SEQUENCE 381 AA; 39199 MW; 97CD8FF5867954EC CRC64;
Query Match	42.1%; Score 228.5; DB 16; Length 381;
Best Local Similarity	48.9%; Pred. No. 3.6e-11;
Matches	45; Conservative 10; Mismatches 36; Indels 1; Gaps
QY	4 CRATVYVNSDWGSGFTATVTVTWTGSRATSCWTVANSGNQTVTWYNTALTQSGASVT 63
Db	279 CAVSYGTVN-VQDQGTADVTVTGAPVDGWLAFVLPDSQGITNANASLTSSGSGVT 337
QY	64 ATNLSTNNVTPQGSSTFFGNSGYSGTNTAPT 95
Db	338 ATGASHNARIAPGSLSFQGTGYGAFAPT 369
RESULT 11	
OB6728	
ID	OB6728 PRELIMINARY; PRT; 973 AA.
AC	OB6728;
DT	01-NOV-1998 (TrEMBLrel. 08, Created)
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Putative secreted cellulase.
GN	SCO6546 OR SC5C7.31C.
OS	Streptomyces coelicolor.

RP SEQUENCE FROM N.A.
RA Vetrivel K.S., Pandian S.K., Chaudary U., Dharmalingam K.:
RT "Purification, cloning and DNA sequence analysis of a chitinase from
RT an overproducing mutant of Streptomyces peucetius defective in
RT daunorubicin biosynthesis.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL: AF206633; AAF43629.1; -;
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR001579; Chitinase_18/2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001223; Glyco_hydro_18.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR ProDom: PD000471; Glyco_hydro_18; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 629 AA; 65410 MW; 11E6214DFE8DC0F8 CRC64;

Query Match 41.3%; Score 224; DB 2; Length 629;
Best Local Similarity 43.9%; Pred. No. 1.4e-10;
Matches 43; Conservative 12; Mismatches 35; Indels 8; Gaps 1;

QY 1 GVGC-----RATVYVNSDMGSGGTATVTNTGSRATSGMTVAWSFGGNGQTVTNYN 52
DB 44 GLACPAEAATATATYTKTSDMGSGFGSWTKVTKTCTTTINSWTVWDFPSPGKVTSAWD 103

QY 53 TALTOSGASVTATNLVNNVVGPGSTTFGNGSYST 90
DB 104 ATVTNSGDHWAKNVGNWGTLPAGTIVSFGNGSGGS 141

RESULT 13
Q08166 PRELIMINARY; PRT; 974 AA.
AC Q08166;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Cellulase 1 precursor (EC 3.2.1.4) (Endoglucanase)
DE (Endo-1,4-beta-glucanase) (Avicelase).
GN Fl.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RA Jung E.D., Lao G., Irwin D., Barr B., Benjamin A., Wilson D.B.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
CC HYDROLASES).
DR EMBL: L20094; AAC06387.1; -;
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR003305; CBM_CenC.
DR InterPro: IPR004197; celd_N.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001701; GH_9.
DR InterPro: IPR000566; Lipocin_cytFABP.
DR InterPro: IPR001230; Prenyl_site.
DR Pfam: PF00553; CBM_2; 1.

DR Pfam: PF02018; CBM_4_9; 1.
DR Pfam: PF02927; celd_N; 1.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 974 CELLULOSE 1.
FT DOMAIN 870 974 CELLULOSE-BINDING.
FT ACT_SITE 673 723 BY SIMILARITY.
FT ACT_SITE 723 732 BY SIMILARITY.
FT ACT_SITE 732 732 BY SIMILARITY.
SQ SEQUENCE 974 AA; 104577 MW; 17FE7330404A83C CRC64;

Query Match 40.5%; Score 220; DB 2; Length 974;
Best Local Similarity 46.7%; Pred. No. 4.7e-10;
Matches 43; Conservative 15; Mismatches 32; Indels 2; Gaps 2;

QY 4 CRATVYVNSDMGSGGTATVTNTGSRATSGMTVAWSFGGNGQTVTNYNNTALTOSGASVT 63
DB 871 CAVTYQTN-DWPGGFTASVTLTNTGSTPMDSWELRFTFPSCQTVSHGWSANWQSGSDVT 929

QY 64 ATNLVNNVVIOP-GOSTTFGNGSYSGTNTAP 94
DB 930 ATSLPWNQSVPPCGGSGVNIQFNGTWCNSNTPK 961

RESULT 14
Q9S675 PRELIMINARY; PRT; 270 AA.
AC Q9S675;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Chitinase (Fragment).
GN CHI.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M145;
RA Kormanec J., Homerova D., Sevcikova B.;
RT "Two-component regulatory system in Streptomyces coelicolor A3(2).";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL: AF102271; AAD16453.1; -;
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_cellose-bind.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001223; Glyco_hydro_18.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR ProDom: PD000471; Glyco_hydro_18; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
KW Hydrolase.
FT NON_TER 270 270
SQ SEQUENCE 270 AA; 27226 MW; D3361F55A258269A CRC64;

Query Match 39.2%; Score 213; DB 2; Length 270;
Best Local Similarity 45.9%; Pred. No. 4.6e-10;
Matches 39; Conservative 13; Mismatches 33; Indels 0; Gaps 0;

QY 6 ATTVVNSDMGSGGTATVTNTGSRATSGMTVAWSFGGNGQTVTNYNNTALTOSGASVTAT 65
DB 36 ATFAKTSMDGTGGGSGWTKVTKTTSLSWTVWDFPPTGKTVTSAMDATVTNSGDHWITAK 95

Result No.	Score	Query Match	Length	DB ID	Description
1	399	73.5	521	17 AAR69927	A. cellulolyticus
2	399	73.5	521	22 AAB48786	Acidothermus cellu
3	399	73.5	521	22 AAB48787	Acidothermus cellu
4	399	73.5	521	22 AAB48788	Acidothermus cellu
5	399	73.5	562	21 AAY69508	Acidothermus cellu
6	399	73.5	562	23 AAU79549	A. cellulolyticus
7	275	50.6	596	17 AAR90715	Thermotstable cellu
8	244.5	45.0	109	18 AAW36180	Cellulomonas fimi
9	244.5	45.0	212	13 AAR20472	Eglin C fusion pro
10	242.5	44.7	467	22 AAB31372	Amino acid sequen

PR 15-JUL-1994; 94US-0276213.

PR 26-SEP-1989; 89US-0412434.
 PR 27-JAN-1992; 92US-0826089.
 PR 21-SEP-1993; 93US-0125115.
 XX
 XX (MIDE) MIDWEST RES INST.
 XX
 XX Himmel ME, Laymon RA, Thomas SR;
 XX
 XX WPI; 1996-105843/11.
 DR N-PSDB; AAT12337.
 DR
 XX New isolated DNA encoding endoglucanase - obtd from Acidothermus
 PT cellulolyticus, used for prodn of the enzyme for use in cellulose
 PT hydrolysis.
 XX
 XX Claim 1: Page 22; 34pp; English.
 XX
 XX Acidothermus cellulolyticus E1 endoglucanase (AAR89927) is useful
 CC for hydrolysing cellulosic biomass to sugars for simultaneous or
 CC subsequent fermentation to ethanol. It shows optimal activity at
 CC 83 deg C. The amino acid sequence of the mature enzyme was deduced
 CC from an E1 endoglucanase gene (AAT12337) cloned from A. cellulolyticus.
 CC Putative signal peptides were also identified (see also AAR89928-29).
 CC Cloning of this gene allows large-scale, low-cost prodn. of
 CC recombinant E1 endoglucanase, using pref. Saccharomyces, Zymomonas
 CC or E.coli hosts.
 XX
 XX Sequence 521 AA;
 SQ
 Query Match 73.5%; Score 399; DB 17; Length 521;
 Best Local Similarity 69.3%; Pred. No. 9.4e-30;
 Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
 QY 1 GVCGRATYVNSDMGSGFTATVTNTGSRATSGWTVAWSFGGNGQTNTYNTALTQSGA 60
 DB 421 GACTASYQVNSDMGNGFTVTVAVTNSGVSATKTWTVSGTITNSWNAAVTONGQ 480
 QY 61 SVTATNLNNYVNIQPGOSTTTFGNGSYSGTNTAPTTLTCTAS 101
 DB 481 SVTARNNSYNNVNIQPGQNTTFGFOASYTGSNAAPTACAAAS 521
 RESULT 2
 AAB48786
 ID AAB48786 standard; protein: 521 AA.
 XX
 XX AAB48786;
 XX
 XX 09-MAR-2001 (first entry)
 XX
 XX Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.
 XX
 XX E1 endoglucanase; glycosyl hydrolase; soluble substrate;
 KW cellulose hydrolysis; ethanol production; fermentation;
 KW mutant; mutein.
 XX
 XX Acidothermus cellulolyticus.
 OS Synthetic.
 OS
 XX WO200070031-A1.
 XX
 XX 23-NOV-2000.
 XX
 XX 19-MAY-2000; 2000WO-US13971.
 XX
 XX 19-MAY-1999; 99US-0134925.
 XX
 XX (MIDE) MIDWEST RES INST.
 XX
 XX Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
 PI Decker SR;
 XX
 XX WPI; 2001-061226/07.
 DR

XX Preparation of glycosyl hydrolase with an increased catalytic activity
 PT on insoluble substrate -
 PT
 XX Claim 16: Page 22-24; 30pp; English.
 XX
 XX The invention relates to a method for making glycosyl hydrolase mutants
 CC with increased catalytic activity with either insoluble or soluble
 CC cellulose substrates relative to the wild-type enzyme. The method for
 CC making a glycosyl hydrolase with increased insoluble substrate catalytic
 CC activity comprises replacing an active site-associated amino glycosyl-
 CC stabilising amino acid with an amino acid that does not strongly bind a
 CC disaccharide product in the active site. Conversely, the method for
 CC making a glycosyl hydrolase with increased soluble substrate catalytic
 CC activity comprises replacing a hydrophobic substrate-binding amino acid
 CC with a positively charged residue. The invention also discloses mutants
 CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)
 CC produced according to the method of the invention. The Y245G mutant
 CC (AAB48788) has improved activity with insoluble substrates, and the W42R
 CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with
 CC soluble substrates. The invention also encompasses DNA encoding these
 CC mutants. The glycosyl hydrolases of the invention are used as catalysts
 CC for cellulose hydrolysis to produce sugars that can be fermented to
 CC produce fuels such as ethanol. The present sequence represents the
 CC Acidothermus cellulolyticus E1 endoglucanase W42R mutant.
 XX
 XX Sequence 521 AA;
 SQ
 Query Match 73.5%; Score 399; DB 22; Length 521;
 Best Local Similarity 69.3%; Pred. No. 9.4e-30;
 Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
 QY 1 GVCGRATYVNSDMGSGFTATVTNTGSRATSGWTVAWSFGGNGQTNTYNTALTQSGA 60
 DB 421 GACTASYQVNSDMGNGFTVTVAVTNSGVSATKTWTVSGTITNSWNAAVTONGQ 480
 QY 61 SVTATNLNNYVNIQPGOSTTTFGNGSYSGTNTAPTTLTCTAS 101
 DB 481 SVTARNNSYNNVNIQPGQNTTFGFOASYTGSNAAPTACAAAS 521
 RESULT 3
 AAB48787
 ID AAB48787 standard; protein: 521 AA.
 XX
 XX AAB48787;
 XX
 XX 09-MAR-2001 (first entry)
 XX
 XX Acidothermus cellulolyticus E1 endoglucanase mutant, Y82R.
 XX
 XX E1 endoglucanase; glycosyl hydrolase; soluble substrate;
 KW cellulose hydrolysis; ethanol production; fermentation;
 KW mutant; mutein.
 XX
 XX Acidothermus cellulolyticus.
 OS Synthetic.
 OS
 XX WO200070031-A1.
 XX
 XX 23-NOV-2000.
 XX
 XX 19-MAY-2000; 2000WO-US13971.
 XX
 XX 19-MAY-1999; 99US-0134925.
 XX
 XX (MIDE) MIDWEST RES INST.
 XX
 XX Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
 PI Decker SR;
 XX
 XX WPI; 2001-061226/07.
 DR

XX PA (MIDE) MIDWEST RES INST.
XX PI Laymon RA, Adney WS, Thomas SR, Himmel ME;
XX PR WPI: 2000-087663/08.
DR N-PSDB: AA25924.
XX Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful
PT for labeling or modifying a cellulose and for purifying or immobilizing
PT a binding domain fusion protein to cellulose .
XX PS Claim 1: Fig 2: 85pp; English.
XX This sequence represents the Acidothermus cellulolyticus E1
CC endoglucanase, which is a beta-1,4-endoglucanase, or endocellulase. The
CC cellulose binding domain (CBD) of E1 endoglucanase, and nucleotides which
CC encode it are specifically claimed. The CBD is believed to be roughly
CC wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge
CC tip is inserted between the microfibrils of the cellulose fibre,
CC disrupting the crystalline structure, and making the cellulose linkages
CC more accessible to the catalytic domain of the E1 endoglucanase. The E1
CC endoglucanase CBD is useful in labelling or modifying the surface of
CC cellulose or other polysaccharides. Such modified cellulose can then be
CC used in textile, pulp, paper, chemical and pharmaceutical industries.
CC CBDs can be used in affinity purification of CBD-fusion proteins, and can
CC also be used to immobilise the CBD-fusion proteins to a cellulose
CC support. CBD-fusion proteins can be used to modify the chemical or
CC physical properties of a cellulose or polysaccharide matrix column and to
CC modify (e.g., roughen or disrupt) a cellulose or polysaccharide fibre.
CC The CBD of the E1 endoglucanase exhibits greater stability at pH 4-8 and
CC has an optimum temperature for stability of 83 degrees Celsius which is
CC not found in CBDs from non-thermophilic organisms.
XX SQ Sequence 562 AA;
Query Match 73.5%; Score 399; DB 21; Length 562;
Best Local Similarity 69.3%; Pred. No. 1e-29;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
QY 1 GVGCRTATVYVNSDVGSGFTATVTNTGSRATSGTVAWSFGGNOTVTNYNTALTQSGA 60
Db 462 GARTASQVNSDVGNGFTTVAVTNSGVSATKTWTVTGFGNQTITNSNNAVTONGQ 521
QY 61 SVTATNLNYYNNVIOPGOSTTTFGNGSYSGTNTAPTLCCTAS 101
Db 522 SVTARNMSYNNVIOPGONTTTFGFOASYTGSNAAPTACAAS 562
RESULT 6
AAU79549
XX AAU79549 standard; Protein; 562 AA.
XX AAU79549;
XX 24-SEP-2002 (first entry)
XX A. cellulolyticus cellulase EI beta-1,4-endoglucanase precursor.
XX Gene; ds; transgenic; plant; lignocellulose; cellulase; ligninase;
KW fermentable sugar; ethanol; fermentation; silage; feed; fuel;
KW Industrial chemical; biodegradation; chloroaromatic;
KW environmental pollutant; EI beta-1,4-endoglucanase; el.
XX Acidothermus cellulolyticus.
XX Key Location/Qualifiers
FH Peptide 1..41
FT /label= Leader_sequence
FT Protein 42..562
FT /label= Mature_cellulase
XX WO200234926-A2.

XX PD 02-MAY-2002.
XX PF 18-OCT-2001: 2001WO-US32538.
XX PR 20-OCT-2000: 2000US-242408P.
XX PA (UNMS) UNIV MICHIGAN STATE.
XX PI Sticklen MB, Dale BE, Magbool S;
XX WPI: 2002-489947/52.
DR N-PSDB: ABK86729.
XX Producing transgenic plants which after harvest degrade lignin and
PT cellulose to fermentable sugars, by mating transgenic plant comprising
PT DNA encoding cellulase with transgenic plant comprising DNA encoding
PT ligninase .
XX Disclosure; Page 96-99; 126pp; English.
XX The invention discloses the production of a transgenic plant which
CC degrades lignocellulose when the plant is grown. It comprises the
CC production of the transgenic plant including cellulase and ligninase by
CC mating a transgenic plant, containing a DNA encoding a cellulase, and a
CC transgenic plant, containing a DNA encoding a ligninase, where both genes
CC are operably linked to a nucleotide sequence encoding a signal peptide
CC which targets the fusion protein to an organelle of the plant,
CC particularly chloroplasts. The method is useful for producing a
CC transgenic plant (e.g. maize) which degrades lignocellulose when the
CC plant is grown to produce a plant material. This material is useful for
CC converting lignocellulose, in a plant material, to fermentable sugars
CC which are then fermented to ethanol. The transgenic plants also provide a
CC plentiful and inexpensive source of fungal or bacterial cellulases and
CC ligninases which can be used in the production of ethanol. They can also
CC be used for pre-treating silage to increase the energy value of
CC lignocellulosic feeds for cows and other ruminant animals, pre-treating
CC lignocellulosic biomass for fermentative conversion to fuels and
CC industrial chemicals, and biodegradation of chloroaromatic environmental
CC pollutants. The protein sequence presented is the A. cellulolyticus
CC cellulase EI beta-1,4-endoglucanase precursor.
XX SQ Sequence 562 AA;
Query Match 73.5%; Score 399; DB 23; Length 562;
Best Local Similarity 69.3%; Pred. No. 1e-29;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
QY 1 GVGCRTATVYVNSDVGSGFTATVTNTGSRATSGTVAWSFGGNOTVTNYNTALTQSGA 60
Db 462 GARTASQVNSDVGNGFTTVAVTNSGVSATKTWTVTGFGNQTITNSNNAVTONGQ 521
QY 61 SVTATNLNYYNNVIOPGOSTTTFGNGSYSGTNTAPTLCCTAS 101
Db 522 SVTARNMSYNNVIOPGONTTTFGFOASYTGSNAAPTACAAS 562
RESULT 7
AAR90715
ID AAR90715 standard; Protein; 596 AA.
XX AAR90715;
AC AAR90715;
XX 06-APR-1996 (first entry)
XX Thermostable cellulase-E3.
XX Cellulase-E3; Thermomonospora fusca; thermostable; plasmid pS24;
KW plasmid pS26; Escherichia coli; cloning; Streptomyces lividans;
KW papain; catalytic domain; cellulose; chitosan; protease resistance;
KW synergism; cellobiohydrolase; beta-glucosidase; saccharification;
KW surfactant additive; paper recycling; delinking; paper refining.

OS Thermomonospora fusca strain YX36.
 XX
 FH Ke/ Location/Qualifiers
 FT Peptide 1..38
 FT /note= "Signal peptide"
 FT Peptide 39..48
 FT /note= "N-terminal peptide (AAR90717)"
 FT Domain 174..596
 FT /note= "Catalytically active domain (AAR90716)"
 XX
 PN W09600281-A1.
 XX
 PD 04-JAN-1996.
 XX
 PF 23-JUN-1995; 95WO-US09069.
 XX
 PR 24-JUN-1994; 94US-0265429.
 XX
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Walker LP, Wilson DB, Zhang S;
 XX
 DR WPI: 1996-068865/07.
 DR N-PSDB; AAT15595.
 XX
 CC Cellulase from Thermomonospora fusca - is thermostable and is useful
 PT in a variety of industrial applications e.g. clarification of fruit
 PT juices, fabric softening, etc
 XX
 PS Claim 1: Page 32-37; 53pp; English.
 XX
 CC The sequence represents thermostable cellulase-E3 (EC-3.2.1.4) from
 CC Thermomonospora fusca, and is encoded by inserts in plasmid pS24
 CC and plasmid pS26, obtained by cloning in Escherichia coli. The
 CC gene may be cloned and expressed in E. coli or Streptomyces
 CC lividans for recombinant cellulase-E3 production. Cellulase-E3,
 CC or a catalytically active polypeptide obtained by papain digestion,
 CC hydrolyses cellulose or chitosan at pH 5-11 and 40-70 deg C, and
 CC has significant activity at over 60 deg C. E3 has higher stability
 CC to proteolysis in culture supernatant than T. fusca cellulases-E2
 CC and -E5, and shows strong synergistic activity when combined with
 CC other cellulases, cellobiohydrolase and optionally beta-glucosidase.
 CC The cellulase and mixtures may be used e.g. in cellulose
 CC saccharification for ethanol production, fruit juice clarification,
 CC as a surfactant additive, or in deinking or refining of recycled
 CC paper.
 XX
 SQ Sequence 596 AA;
 Query Match 50.6%; Score 275; DB 17; Length 596;
 Best Local Similarity 52.4%; Pred. No. 6.7e-18;
 Matches 55; Conservative 10; Mismatches 32; Indels 8; Gaps 3;
 QY 3 GCRATYVNSDWSGSGFTATVTVTGSRATSGWTVAVSFGNGQTNTALTQSGASV 62
 DB 40 GCSVDVTVNS-WGTGFTANTVTINLGS-AINGWTLEWDFNGQOVTNLWNGTYTQSGQHV 97
 QY 63 TATNLNYSNNVIOQGQSTTFGNGSYSGTNTAPT-----LTCTAS 101
 DB 98 SVSNAPYNASIPANGVTEGFNGSYSGNSNDIPSSFKLNGVTCDGS 142
 RESULT 8
 AAW36180
 ID AAW36180 standard; Protein: 109 AA.
 XX
 AC AAW36180;
 XX
 DT 15-APR-1998 (first entry)
 XX
 DE Cellulomonas fimi Cex cellulose binding domain.
 XX
 KW Cellulose binding domain; Cex; cellulase; fusion protein; growth factor;

KW immobilisation; cell surface; steel factor; haematopoietic cell; blood;
 KW pluripotent cell; stem cell; bone marrow; megakaryocyte.
 OS Cellulomonas fimi.
 XX
 PN W09726358-A1.
 XX
 PD 24-JUL-1997.
 XX
 PF 16-JAN-1997; 97WO-CA00033.
 XX
 PR 16-JAN-1996; 96US-0585585.
 XX
 PA (UYBR-) UNIV BRITISH COLUMBIA.
 XX
 PI Alimonti J, Doheny JG, Humphries KR, Jervis E, Kilburn DG;
 XX
 DR WPI: 1997-385344/35.
 XX
 CC Composition for expanding or isolating cells dependent on specific
 PT growth factor - comprises the growth factor conjugated to substrate
 PT binding part of polysaccharidase for immobilisation plus a growth
 PT factor receptor, optionally on a cell surface
 XX
 PS Disclosure; Fig 1; 100pp; English.
 XX
 CC This is the amino acid sequence of the cellulose binding domain from the
 CC Cex cellulase from Cellulomonas fimi. The domain can be used to generate
 CC a fusion protein, especially for immobilisation of a growth factor (GF)
 CC on e.g. a cell surface. The method comprises binding the cellulose
 CC binding domain of the cellulase to the growth factor, immobilising the
 CC composition and releasing the growth factor from the cellulose binding
 CC domain, e.g. by cleavage with a protease. The compositions are used for
 CC in vitro expansion of GF-dependent cells (especially where dependent on
 CC steel factor (SF), particularly haematopoietic cells (pluripotent or stem
 CC cells from blood or bone marrow) or megakaryocytes); and to produce a
 CC population enriched in cells having a GF-receptor (e.g. rare cells such
 CC as pluripotent stem cells).
 XX
 SQ Sequence 109 AA;
 Query Match 45.0%; Score 244.5; DB 18; Length 109;
 Best Local Similarity 49.5%; Pred. No. 7e-16;
 Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;
 QY 3 GCRATYVNSDWSGSGFTATVTVTGSRATSGWTVAVSFGNGQTNTALTQSGASV 62
 DB 5 GCQVLCGVN-QWNTGFTANTVTNKNTSSAPVDGWTLTFSFGQGVTOAWSSTVTQSGSAV 63
 QY 63 TATNLNYSNNVIOQGQSTTFGNGSYSGTNTAPT 95
 DB 64 TVRNAPWNGSIPAGGTAGQFGNGSHTGTNAAPT 96
 RESULT 9
 AAR20472
 ID AAR20472 standard; Protein: 212 AA.
 XX
 AC AAR20472;
 XX
 DT 15-APR-1992 (first entry)
 XX
 DE Eglin C fusion protein with KEX2/KEX1 site.
 XX
 KW Cellulose binding; KEX2/KEX1; yscF; ysc alpha; endoprotease; Exg;
 KW carboxypeptidase; cex.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS Cellulomonas fimi.
 XX
 FH Key Location/Qualifiers
 FT Region 1..71

```
FT FT /label= eglin_C
FT 72..79
FT /label= linker
FT 80..212
FT /label= cellulose_binding_domain
FT /note= "from Cellulomonas fimi Exg protein "
FT 72..73
FT /label= KEX2/KEX1_recognition
XX EP467839-A.
PN
XX
XX 22-JAN-1992.
XX
XX 09-JUL-1991; 91EP-0810543.
XX
XX 18-JUL-1990; 90GB-0015825.
XX
XX (CIBA ) CIBA GEIGY AG.
XX
XX Helm J, Seeboth P, Takabayashi K:
XX WPI 1992-026641/04.
XX
XX In vitro processing of fusion protein - using yeast endo:protease
PT ysc F and carboxy:peptidase ysc alpha, for prodn. of biologically
PT active mature protein
XX
XX Disclosure; Page 30; 35pp; English.
XX
XX The sequence (SEQ ID NO: 4) is encoded by an insert from
CC PDP34GAPDH- eglinex-1, a vector constructed for the expression of
CC the fusion protein in Saccharomyces cerevisiae. The protein has a
CC formula P-L-T, in which P is the eglin C polypeptide (Rink H.
CC et al. 1984 Nucl. Acids Res. 12 6369-6387), L a linker incorporating
CC the KEX2/KEX1 recognition site, and T the cellulose binding domain
CC of the C. fimi Exg protein (including an extra proline) encoded by
CC the cex gene (Neill G.O. et al. 1986 Gene 44 325-330). On expression
CC the crude fusion protein is purified by cellulose affinity
CC chromatography. Fractions contg. the protein are then treated with
CC sol. yscF resulting in the cleavage prod. eglin C-Lys-Arg which is
CC then digested with sol. ysc alpha to remove the Lys-Arg tail.
CC The obtd. eglin C is identical in structure to natural eglin C
CC as evidenced by HPLC.
CC See also AAR20469-71.
XX
XX SQ Sequence 212 AA:
Query Match 45.0%; Score 244.5; DB 13; Length 212;
Best Local Similarity 49.5%; Pred. No. 1.5e-15;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;
OY 3 GCRATYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGNGNOTVTNWNLTALTSQGSV 62
Db 109 GCQVLGCVN-QWNTGFTANVTYKNTSAPVDGWTLTFFSPSGQOVTOAMSSTVTQSGSAV 167
OY 63 TATNLSYNNVIOPCQSTTFGNGSYSGTNAPT 95
Db 168 TVRNAPMNGSIPAGGTAQFGNGSHTGTNAAPT 200
RESULT 10
AAB31372
ID AAB31372 standard; Protein; 467 AA.
XX
XX AAB31372;
XX
XX 20-APR-2001 (first entry)
XX
XX Amino acid sequence of protein L/CBD cex/ER retaining peptide fusion.
XX Protein production: food processing; protein antibiotic; feed enzyme;
KW protein L: CBD cex protein; cell signal peptide.
XX
```

```
OS Synthetic.
XX
XX PN WO20007174-A1.
XX
XX PD 21-DEC-2000.
XX
XX PF 07-JUN-2000; 2000WO-IL00330.
XX
XX PR 10-JUN-1999; 99US-0329234.
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Shani Z, Shoseyov O;
XX
XX WPI: 2001-112219/12.
DR N-PSDB; AAF24730.
XX
XX Expressing and isolating recombinant protein in a plant, useful for
PT producing large quantities of recombinant proteins, by expressing a
PT fusion protein including a cellulose binding peptide fused to a
PT recombinant protein -
XX
XX Example: Fig 2a; 87pp; English.
XX
XX The specification describes a method for expressing and isolating a
CC recombinant protein in a plant. The method comprising expressing a
CC fusion protein including the recombinant protein and a cellulose
CC binding peptide fused to it, where the fusion protein is
CC compartmentalised and sequestered within plant cells, for obtaining large
CC tissue or cultured plant cells. The method is useful for obtaining large
CC quantities of the recombinant proteins and protein products in a simple
CC and cost-effective manner. Recombinant proteins may be used commercially,
CC such as in the food processing industry, e.g. glucoamylases and glucose
CC isomerases are used for converting starch to high fructose corn syrup,
CC pectin hydrolysis in food industry, lipases for cleaving ester linkage
CC in triglycerides, and for effluent treatment. The recombinant proteins
CC may further be used to produce protein antibiotics, which can be used
CC in healing processes, and to produce animal feed enzymes. The present
CC sequence represents a fusion protein of the invention, and comprises a
CC fusion of a cell signal peptide, protein L, CBD cex, and an endoplasmic
CC reticulum retaining peptide.
XX
XX SQ Sequence 467 AA:
Query Match 44.7%; Score 242.5; DB 22; Length 467;
Best Local Similarity 49.5%; Pred. No. 6.2e-15;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;
OY 3 GCRATYVNSDMGSGFTATVTNTGSRATSGMTVAWSFGNGNOTVTNWNLTALTSQGSV 62
Db 360 GCQVLGCVN-QWNTGFTAQVTVKNTSAPVDGWTLTFFSPSGQOVTOAMSSTVTQSGSAV 418
OY 63 TATNLSYNNVIOPCQSTTFGNGSYSGTNAPT 95
Db 419 TVRNAPMNGNIPAGGTAQFGNGSHTGTNAAPT 451
RESULT 11
AAB31373
ID AAB31373 standard; Protein; 509 AA.
XX
XX AAB31373;
XX
XX 20-APR-2001 (first entry)
XX
XX Amino acid sequence of Sig/CBD cex/ER retaining peptide fusion.
XX Protein production: food processing; protein antibiotic; feed enzyme;
KW protein L: CBD cex protein; cell signal peptide.
XX
```

OS Synthetic.
 PN WO2000077174-A1.
 XX 21-DEC-2000.
 XX 07-JUN-2000; 2000WO-IL00330.
 XX 10-JUN-1999; 99US-0329234.
 XX (CBDT-) CBD TECHNOLOGIES LTD.
 PA (VISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX Shani Z, Shoseyov O;
 XX WPI; 2001-112219/12.
 DR N-PSDB; AAF24735.
 XX Expressing and isolating recombinant protein in a plant, useful for
 PT producing large quantities of recombinant proteins, by expressing a
 PT fusion protein including a cellulose binding peptide fused to a
 PT recombinant protein -
 XX Example: Fig 3a; 87pp; English.
 PS The specification describes a method for expressing and isolating
 CC a recombinant protein in a plant. The method comprising expressing a
 CC fusion protein including the recombinant protein and a cellulose
 CC binding peptide fused to it, where the fusion protein is
 CC compartmentalised and sequestered within plant cells, plant derived
 CC tissue or cultured plant cells. The method is useful for obtaining large
 CC quantities of the recombinant proteins and protein products in a simple
 CC and cost-effective manner. Recombinant proteins may be used commercially,
 CC such as in the food processing industry, e.g. glucoamylases and glucose
 CC isomerases are used for converting starch to high fructose corn syrup,
 CC proteinases for the hydrolysis of high molecular weight proteins and in
 CC manufacturing leather or alcoholic beverages, pectinesterases for
 CC pectin hydrolysis in food industry, lipases for cleaving ester linkage
 CC in triglycerides, and for effluent treatment. The recombinant proteins
 CC may further be used to produce protein antibiotics, which can be used
 CC in healing processes, and to produce animal feed enzymes. The present
 CC sequence represents a fusion protein of the invention, and comprises a
 CC fusion of a cell signal peptide, Sig (not specified), CBD cex, and an
 CC endoplasmic reticulum retaining peptide.
 XX
 XX Sequence 509 AA;
 S0
 Query Match 44.7%; Score 242.5; DB 22; Length 509;
 Best Local Similarity 49.5%; Pred. No. 6.0e-15;
 Matches 46; Conservative 14; Mismatches 32; Indels 1; Gaps 1;
 QY 3 GCRTATVYVNSDWGSGFTATVTVTGTGSRATSGWTVVAVSFGGNGQTVTVYNTALTQSGASV 62
 DB 41 GCQVLWGVN-QWNTGFTAQVTVKNTGAPVDGWTLTFTSPSGQQVTOAWSSTVTVQSGSAV 99
 QY 63 TATNLNSVNNVIOQCGSTTFGNGSYSGTNTAPT 95
 DB 100 TVRNAPWNGNIPAGGTAQFGQSGHTGTNAAPT 132
 RESULT 12
 ABB06928
 ID ABB06928 standard; Protein; 518 AA.
 XX ABB06928;
 AC
 XX 18-JUN-2002 (first entry)
 DT
 XX Micromonospora carbonacea everninomicin locus protein ORF 47.
 DE
 XX Micromonospora carbonacea; antibiotic; everninomicin; biosynthesis;
 KW gene cluster; genetic manipulation; contig.
 XX

OS Micromonospora carbonacea.
 XX WO200155180-A2.
 PN 02-AUG-2001.
 XX 29-JAN-2001; 2001WO-CA00128.
 PF 27-JAN-2000; 2000US-0177711.
 XX (ECOP-) ECOPIA BIOSCIENCES INC.
 PA (FARN/) FARNET C.
 XX Staffa A, Zazopoulos E, Mercure S, Nowacki P;
 PI WPI; 2001-476185/51.
 XX N-PSDB; ABL50562.
 DR Novel isolated gene cluster encoding polypeptides involved in
 XX everninomicin biosynthesis useful for construction of everninomicin
 PT overproducing strains, and to allow chemical modifications of
 PT everninomicin to enhance certain properties -
 XX Claim 15; Fig 1; 181pp; English.
 PS ABL50555 to ABL50562 represent contigs 1 to 8 from the Micromonospora
 CC carbonacea everninomicin biosynthetic locus gene cluster. The contigs
 CC encode the protein sequences designated ORF (open reading frame) 1 to
 CC 49, given in ABB06881 to ABB06930. The gene cluster is useful for the
 CC construction of the everninomicin antibiotic in overproducing strains,
 CC and to allow chemical modifications of everninomicin to enhance certain
 CC properties via genetic manipulation or combinatorial biosynthesis. The
 CC gene cluster can be used to produce genetic systems and genes encoding
 CC novel enzyme activities, and avoid the problems of low yield and quality
 CC of everninomicins produced by chemical synthesis.
 XX
 XX Sequence 518 AA;
 S0
 Query Match 42.5%; Score 231; DB 22; Length 518;
 Best Local Similarity 45.3%; Pred. No. 8.6e-14;
 Matches 48; Conservative 13; Mismatches 37; Indels 8; Gaps 3;
 QY 3 GCRTATVYVNSDWGSGFTATVTVTGTGSRATSGWTVVAVSFGGNGQTVTVYNTALTQSGAS 61
 DB 34 GCSVAYTVQSQWGTGSGNVAITNLGS-ALGTWTLTDFPTSGQQVTOAWSATWSQSGTS 92
 QY 62 VTATNLNSVNNVIOQCGSTTFGNGSYSGTNTAPT-----LTCTAS 101
 DB 93 VSAASLSWNGSLGTGGTGTITGFGWSGSGNPVPKSFALNGTTCTGSG 138
 RESULT 13
 AAY06368
 ID AAY06368 standard; Protein; 381 AA.
 XX AAY06368;
 AC
 XX 06-SEP-1999 (first entry)
 DT
 XX Streptomyces lividans EGI11-like cellulase.
 DE
 XX Cellulase; endoglucanase; EGI11; textile; feed additive; baking;
 KW food processing; grain wet milling; pulp; paper.
 XX Streptomyces lividans.
 OS
 XX WO9931255-A2.
 PN 24-JUN-1999.
 XX 14-DEC-1998; 98WO-US26552.
 PF 16-DEC-1997; 97US-0991720.
 XX

XX (GEMV) GENENCOR INT INC.
 XX PA Bower BS, Fowler T, Phillips JI;
 XX PI WPI; 1999-395187/33.
 XX DR EGIII like cellulase
 XX PT
 XX Example; Fig 6; 47pp; English.
 XX The present polypeptide represents a full-length sequence of a
 CC novel EGIII-like cellulase of Streptomyces lividans (CeIB). It was
 CC deduced from a gene sequence isolated from genomic DNA using PCR
 CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
 CC of Trichoderma reesei EGIII cellulase and related enzymes. PCR
 CC has been used to identify novel EGIII-like enzymes, including the
 CC present protein, from bacterial and fungal sources (see AAY06331-70).
 CC The sequence shows homology to T. reesei EGIII (see AAY06330). Also
 CC provided by the invention are vectors, host cells and methods
 CC for the recombinant production of such enzymes, which can be used
 CC in the treatment of cellulose-containing textiles, as feed
 CC additives, in the treatment of wood pulp, in the reduction of
 CC biomass to glucose, in the stone washing of indigo dyed denim, or
 CC as laundry detergent components (all claimed).
 XX Sequence 381 AA;
 SQ

Query Match 42.1%; Score 228.5; DB 20; Length 381;
 Best Local Similarity 48.9%; Pred. No. 1e-13;
 Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;
 QY 4 CRATYVNSDMGSGFTATVTVTNGSRATSGWTVAWSFGNGQVTNTYNTALTOSGASVT 63
 DB 279 CAVSYGTNV-WQDGFADVTVTGTAPVDCWQLAFTLPSCQRITNAWNASLTPSSGSVT 337
 QY 64 ATNLSYNNVIOGCGSTTFGFGNGSYSGTNTAPT 95
 DB 338 ATGASHNARIAPGGSLSFGFGTYGGAFAEPT 369

RESULT 14
 AABI4881
 ID AABI4881 standard; Protein: 381 AA.
 AC AABI4881;
 XX 21-NOV-2000 (first entry)
 XX Rhodothermus marinus EGIII-like cellulase.
 DE Rhodothermus marinus; Trichoderma reesei; endoglucanase III; EGIII;
 KW cellulase; mutant; enzyme stability; textile treatment;
 KW wood pulp treatment; feed additive; detergent.
 XX Rhodothermus marinus.
 OS
 XX WO200037614-A2.
 XX
 PD -J-JUN-2000.
 PI 12-NOV-1999; 99WO-US26704.
 XX
 XX 10-NOV-1998; 98US-0216295.
 XX (GEMV) GENENCOR INT INC.
 XX Mitchellson C, Wendt DJ;
 XX WPI; 2000-482483/42.
 XX Novel endoglucanase III or endoglucanase III-like cellulase useful for
 PT treating textiles and wood pulp comprises a substitution or deletion at

PT specified positions in the wild form of endoglucanase III -
 XX Example 1; Fig 3; 52pp; English.
 XX The present sequence is a cellulase related to endoglucanase III (EGIII)
 CC from Trichoderma reesei. EGIII-like genes were isolated from genomic DNA
 CC libraries constructed from various microorganisms by PCR. The isolated
 CC genes showed significant homology to EGIII from T. reesei. Certain
 CC substitution and deletion mutations have been incorporated into EGIII and
 CC EGIII-like cellulases to produce variant enzymes with improved stability,
 CC e.g. increased resistance to temperature stress. The mutants may be used
 CC in textile and wood pulp treatment, as a feed additive, and for reducing
 CC biomass to glucose. They are also useful for stonewashing or indigo dyed
 CC denim and as an agent in laundry and dish detergents.
 XX Sequence 381 AA;
 SQ

Query Match 42.1%; Score 228.5; DB 21; Length 381;
 Best Local Similarity 48.9%; Pred. No. 1e-13;
 Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;
 QY 4 CRATYVNSDMGSGFTATVTVTNGSRATSGWTVAWSFGNGQVTNTYNTALTOSGASVT 63
 DB 279 CAVSYGTNV-WQDGFADVTVTGTAPVDCWQLAFTLPSCQRITNAWNASLTPSSGSVT 337
 QY 64 ATNLSYNNVIOGCGSTTFGFGNGSYSGTNTAPT 95
 DB 338 ATGASHNARIAPGGSLSFGFGTYGGAFAEPT 369

RESULT 15
 AAU77433
 ID AAU77433 standard; Protein: 381 AA.
 XX AAU77433;
 XX 05-JUN-2002 (first entry)
 XX Streptomyces lividans CeIB EGIII-like cellulase.
 DE Streptomyces lividans CeIB EGIII-like cellulase.
 KW Endoglucanase III-like cellulase; EGIII-like;
 KW cellulose containing textile; enzyme.
 XX Streptomyces lividans CeIB.
 OS
 XX WO200212464-A2.
 XX
 PD 14-FEB-2002.
 XX 31-JUL-2001; 2001WO-US23989.
 XX 04-AUG-2000; 2000US-0632426.
 XX (GEMV) GENENCOR INT INC.
 XX Mitchellson C, Ropp TH, Swanson BA;
 XX WPI; 2002-241750/29.
 XX Novel endoglucanase III (EGIII)-like cellulase variant comprising
 PT substitution/deletion at positions corresponding to specific residues
 PT in EGIII from Trichoderma reesei, useful for treating cellulose
 PT containing textile -
 XX Example 1; Fig 3; 41pp; English.
 XX The present invention relates to novel endoglucanase III (EGIII)-like
 CC cellulase variants which comprise a substitution or deletion at a
 CC position corresponding to one or more of residues W79, M154 and/or
 CC M118 in mature EGIII from the fungus, Trichoderma reesei. The variants
 CC are useful in the treatment of a cellulose containing textile. By
 CC substituting other amino acids for the native methionines at
 CC positions 79, 118 and 154 (sites where oxidation of the enzyme takes

CC place) in EGIII from T. reesei, oxidatively more stable enzymes are
 CC obtained. The present sequence represents Streptomyces lividans CelB
 CC EGIII-like cellulase.

XX
 SQ Sequence 381 AA:

Query Match 42.1%; Score 228.5; DB 23; Length 381;
 Best Local Similarity 48.9%; Pred. No. 1e-13;
 Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;

Oy 4 CRATVVNSDNGSGFTATVTNTCSRATSGTVAWSFGGNOTVTNYWNTALTQSGASVT 63

Db 279 CAVSYGTNV-WQDGTADVTNTGTAPVDCWQLAFTLPSCORITNAWNASLTPSSGSVT 337

Oy 64 ATNLSYNNVIOPGQSTTFGNGSYSGTNTAPT 95

Db 338 ATGASHNARIAPGGSLSFGFGTYCGAFAPT 369

Search completed: November 13, 2002, 11:51:41
 Job time : 10.0697 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:49:46 ; Search time 3.1197 Seconds
(without alignments)
952.482 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVCCRTATVYVNDWGSFGTA.....GFNGSYSGTNTAPLTCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	399	73.5	521	1	US-08-276-213-3
2	244.5	45.0	108	2	US-08-585-585A-2
3	244.5	45.0	108	2	US-08-249-037C-2
4	244.5	45.0	108	2	US-08-788-622B-2
5	244.5	45.0	108	3	US-08-788-621B-2
6	240.5	44.3	104	2	US-08-585-585A-3
7	240.5	44.3	104	2	US-08-249-037C-3
8	240.5	44.3	104	2	US-08-788-622B-3
9	240.5	44.3	104	3	US-08-788-621B-3
10	230.5	42.4	106	2	US-08-585-585A-1
11	230.5	42.4	106	2	US-08-249-037C-4
12	230.5	42.4	106	2	US-08-788-622B-4
13	230.5	42.4	106	3	US-08-788-621B-4
14	228.5	42.1	381	4	US-09-216-295-22
15	213.5	39.3	490	4	US-09-109-841-2
16	212.5	39.1	109	2	US-08-585-585A-1
17	212.5	39.1	109	2	US-08-249-037C-1
18	212.5	39.1	109	2	US-08-788-622B-1
19	212.5	39.1	109	3	US-08-788-621B-1
20	212.5	39.1	631	3	US-08-814-052-8
21	212.5	39.1	631	3	US-08-812-829-8
22	184.5	34.0	592	1	US-08-217-327-8
23	184	33.9	103	2	US-08-585-585A-5
24	184	33.9	103	2	US-08-249-037C-5
25	184	33.9	103	2	US-08-788-622B-5
26	184	33.9	103	3	US-08-788-621B-5
27	184	33.9	104	2	US-08-585-585A-6

28	184	33.9	104	2	US-08-249-037C-6	Sequence 6, Appli
29	184	33.9	104	2	US-08-788-622B-6	Sequence 6, Appli
30	184	33.9	104	3	US-08-788-621B-6	Sequence 6, Appli
31	179	33.0	430	2	US-08-924-440-2	Sequence 2, Appli
32	162.5	29.9	371	4	US-09-104-308-1	Sequence 1, Appli
33	162.5	29.9	371	4	US-09-321-981-1	Sequence 1, Appli
34	161.5	29.7	386	4	US-09-321-981-5	Sequence 5, Appli
35	122.5	22.6	97	2	US-08-585-585A-7	Sequence 7, Appli
36	122.5	22.6	97	2	US-08-249-037C-7	Sequence 7, Appli
37	122.5	22.6	97	2	US-08-788-622B-7	Sequence 7, Appli
38	122.5	22.6	97	3	US-08-788-621B-7	Sequence 7, Appli
39	116.5	21.5	427	2	US-08-282-197C-53	Sequence 53, Appli
40	94.5	17.4	312	4	US-09-216-295-21	Sequence 21, Appli
41	93	17.1	1026	1	US-08-194-290-7	Sequence 7, Appli
42	93	17.1	1026	2	US-08-614-377A-7	Sequence 7, Appli
43	93	17.1	1026	4	US-09-142-648B-7	Sequence 7, Appli
44	89.5	16.5	221	1	US-08-032-848C-13	Sequence 13, Appli
45	89.5	16.5	259	4	US-09-216-295-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-276-213-3
; Sequence 3, Application US/08276213
; Patent No. 5536655
; GENERAL INFORMATION:
; APPLICANT: Thomas, Steven
; APPLICANT: Laymon, Robert
; APPLICANT: Himmel, Michael
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/276,213
; APPLICATION NUMBER: US/08/276,213
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)231-1000
; TELEFAX: (303)231-1098
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-276-213-3

Query Match 73.5% Score 399; DB 1; Length 521;
Best Local Similarity 69.3%; Pred. No. 1.5e-31;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Gaps 0;

Oy 1 GVCCRTATVYVNDWGSFGTATVTNTCSRATSGTNTVNTALTSQA 60

```
Db 421 GARTASYVNSDWGNGFTVTVAVTNSGVSATKTWTVSFTGGNQTITNSWNAAVTQNG 480
QY 61 SVTATNLNYSNNVOPGOSTTFFGNGSGTNTAPTLCCTAS 101
Db 481 SVTARNMSYNNVOPGONTTFFGQASVTGNSNAAPTACAAS 521

RESULT 2
US-08-585-585A-2
: Sequence 2, Application US/08585585A
: Patent No. 5874308
: GENERAL INFORMATION:
: APPLICANT: Kilburn, Douglas G.
: APPLICANT: Humphries, Keith R.
: APPLICANT: Doheny, James G.
: APPLICANT: Jervis, Eric
: APPLICANT: Allmonti, Jude
: TITLE OF INVENTION: Compositions and methods for modulating
: cell proliferation using growth factor-polysaccharide
: TITLE OF INVENTION: binding fusion proteins
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rae-Venter Law Group
: STREET: Box 60039
: CITY: Palo Alto
: STATE: CA
: COUNTRY: U.S.
: ZIP: 94306
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/585,585A
: FILING DATE: 16-JAN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Kung, Viola T.
: REGISTRATION NUMBER: 41,131
: REFERENCE/DOCKET NUMBER: CBOT.016.000S
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650)328-4400
: TELEFAX: (650)328-4477
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 108 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
US-08-585-585A-2

Query Match 45.0%; Score 244.5; DB 2; Length 108;
Best Local Similarity 49.5%; Pred. No. 3.5e-17;
Matches 46; Conservative 14; Mismatches 32; Indels 1; Caps 1;

QY 3 GCRATVYVNSDWGSGFTATVTVTNTGSRATSGWTVAMSGNGTNTVNTALTOSGASV 62
Db 5 GCOVLGCVN-QWNTGFTANTVTNTSSAPVDGWTLTFFSPSGQOVTOAWSSTVTQSGSAV 63

QY 63 TATNLNYSNNVOPGOSTTFFGNGSGTNTAPT 95
Db 64 TVRNAPWNGSIPAGGTAQFGFNGSHTGTNAAPT 96

RESULT 3
US-08-249-037C-2
: Sequence 2, Application US/08249037C
: Patent No. 5928917
: GENERAL INFORMATION:
: APPLICANT: Kilburn, Douglas G.

Db 421 GARTASYVNSDWGNGFTVTVAVTNSGVSATKTWTVSFTGGNQTITNSWNAAVTQNG 480
QY 61 SVTATNLNYSNNVOPGOSTTFFGNGSGTNTAPTLCCTAS 101
Db 481 SVTARNMSYNNVOPGONTTFFGQASVTGNSNAAPTACAAS 521

RESULT 4
US-08-788-622B-2
: Sequence 2, Application US/08788622B
: Patent No. 5962289
: GENERAL INFORMATION:
: APPLICANT: Kilburn, Douglas G.
: APPLICANT: Miller, Robert C.
: APPLICANT: Warren, Richard A.J.
: APPLICANT: Gilkes, Neil R.
: TITLE OF INVENTION: Polysaccharide binding fusion proteins
: TITLE OF INVENTION: and conjugates
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
```


Query Match 44.3%; Score 240.5; DB 2; Length 104;
Best Local Similarity 45.5%; Pred. No. 8.3e-17;
Matches 46: Conservative 18; Mismatches 36; Indels 1

Qy	1	GVGCRATVYVNSDNGSGFTATVTVNTGSRATSGHTVAMSGGNOTVTVNNTAI	105GA
Db	5 <td>GRACEATVYNQPPGQAEVTKNTGSSP <td>101NGMTVQLMNGDLSTSGS</td> </td>	GRACEATVYNQPPGQAEVTKNTGSSP <td>101NGMTVQLMNGDLSTSGS</td>	101NGMTVQLMNGDLSTSGS
Qy	61 <td>SVATNLSSYNHVIQGGSTTFGFGNSGNTAPTLCTAS</td> <td>101</td>	SVATNLSSYNHVIQGGSTTFGFGNSGNTAPTLCTAS	101
Db	65 <td>NVTVRNYSWNGNVPAGSGTSPFGFLCSGTG</td> <td>104QLSSSITCSAS</td>	NVTVRNYSWNGNVPAGSGTSPFGFLCSGTG	104QLSSSITCSAS

RESULT 8
US-08-788-622B-3
: Sequence 3, Application US/08788622B
: Patent No. 5962289
: GENERAL INFORMATION:
: APPLICANT: Kilburn, Douglas G.
: APPLICANT: Miller, Robert C.
: APPLICANT: Warren, Richard A.J.
: APPLICANT: Gilkes, Neil R.
: TITLE OF INVENTION: Polysaccharide binding fusion proteins
: TITLE OF INVENTION: and conjugates
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rae-Venter Law Group, P.C.
: STREET: P.O.Box 60039
: CITY: Palo Alto
: STATE: CA
: COUNTRY: U.S.
: ZIP: 94306
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08788,622B
: FILING DATE: January 23, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/249,037
: FILING DATE: 24-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/865,095
: FILING DATE: 08-APR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/603,987
: FILING DATE: 25-OCT-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/216,794
: FILING DATE: 08-JUL-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Kung, Viola T.
: REGISTRATION NUMBER: 41,131
: REFERENCE/DOCKET NUMBER: CBDT.002.06US
: TELECOMMUNICATION INFORMATION:

APPLICANT: Warren, Richard A.J.
APPLICANT: Gilkes, Neil R.
TITLE OF INVENTION: Polysaccharide binding fusion proteins
TITLE OF INVENTION: and conjugates
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rae-Venter Law Group, P.C.
STREET: P.O.Box 60039
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/788,621B
FILING DATE: January 23, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/249,037
FILING DATE: 24-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/865,095
FILING DATE: 08-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,987
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/216,794
FILING DATE: 08-JUL-1988
ATTORNEY/AGENT INFORMATION:
NAME: Kung, Viola T.
REGISTRATION NUMBER: 41,131
REFERENCE/DOCKET NUMBER: CBDT.002.050S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)328-4400
TELEFAX: (650)328-4477
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-788-621B-4

Query Match 42.4%; Score 230.5; DB 3; Length 106;
Best Local Similarity 43.1%; Pred. No. 8.2e-16;
Matches 44; Conservative 18; Mismatches 33; Indels 7; Gaps 2;
Qy 4 CRATYVYVNSDMGSGFTATVTNTGSRATSGWTVAWSFGNGQTVTNYWNTALTOSGASVT 63
Db 4 CKVEYNASS-WNTGTASVRYTNTGTTALNGWTLTFEPANGQTVOQGWSADWSOGTTVT 62
Qy 64 ATNLNNVNIOPQSGSTTFGNGSYSGTNTAPT-----LTCT 99
Db 63 AKNAWNGSLAAGTVDIGFNGAHNGTNNKPASFTLNCATCT 104

RESULT 14
US-09-216-295-22
Sequence 22, Application US/09216295
Patent No. 6268328
GENERAL INFORMATION:
APPLICANT: Mitchinson, Colin
APPLICANT: Wendt, Dan J.
TITLE OF INVENTION: No. 6268328el Variant EGIII-Like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 41

SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 22
LENGTH: 381
TYPE: PRT
ORGANISM: Streptomyces lividans CelB
US-09-216-295-22

Query Match 42.1%; Score 228.5; DB 4; Length 381;
Best Local Similarity 48.9%; Pred. No. 5.8e-15;
Matches 45; Conservative 10; Mismatches 36; Indels 1; Gaps 1;
Qy 4 CRATYVYVNSDMGSGFTATVTNTGSRATSGWTVAWSFGNGQTVTNYWNTALTOSGASVT 63
Db 279 CAVSYGTNV-WODGFTADVTNTGTAPVDGWLAFITLPSGORITNAMNASLTPSSGCVT 337
Qy 64 ATNLNNVNIOPQSGSTTFGNGSYSGTNTAPT 95
Db 338 ATGASHNARIAPGGSLSFGFQGTGCAFAEPT 369

RESULT 15
US-09-109-841-2
Sequence 2, Application US/09109841
Patent No. 6207436
GENERAL INFORMATION:
APPLICANT: Bjornvad, Mads E.
APPLICANT: Hatakeyama, Mariko
APPLICANT: Schulein, Martin
APPLICANT: Nielsen, Jack B.
TITLE OF INVENTION: Endo-Beta-1, 4-Glucanases From
FILE REFERENCE: 5195.200-US
CURRENT APPLICATION NUMBER: US/09/109,841
CURRENT FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: 0812/97
EARLIER FILING DATE: 1997-07-04
EARLIER APPLICATION NUMBER: 0846/97
EARLIER FILING DATE: 1997-07-11
EARLIER APPLICATION NUMBER: 60/053,506
EARLIER FILING DATE: 1997-07-23
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 490
TYPE: PRT
ORGANISM: Saccharothrix australiensis
US-09-109-841-2

Query Match 39.3%; Score 213.5; DB 4; Length 490;
Best Local Similarity 44.1%; Pred. No. 2.3e-13;
Matches 41; Conservative 11; Mismatches 40; Indels 1; Gaps 1;
Qy 3 GCRTYVYVNSDMGSGFTATVTNTGSRATSGWTVAWSFGNGQTVTNYWNTALTOSGASV 62
Db 42 CCRVDYAVTSQWPGCGAAVVTNLGD-PLSSWELSWTFPPDQGVQOLWNGVHSTSGSV 100
Qy 63 TATNLNNVNIOPQSGSTTFGNGSYSGTNTAPT 95
Db 101 TVKEMSWNGSVGTNASVQVGFNGSHNGANNAPT 133

Search completed: November 13, 2002, 11:56:06
Job time : 3.11997 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2002, 11:55:11 : Search time 1.95905 Seconds
(without alignments)
776.473 Million cell updates/sec

Title: US-09-917-378-5

Perfect score: 543

Sequence: 1 GVGCRTATVYVNSDMGSGFTA.....GFNGSYSGTNTAPTTLCTAS 101

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pdb:*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pdb:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pdb:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pdb:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pdb:*

6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pdb:*

7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pdb:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pdb:*

9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pdb:*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pdb:*

11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pdb:*

12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pdb:*

13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pdb:*

14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pdb:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	399	73.5	562	10	US-09-981-900B-5
2	162.5	29.9	371	10	US-09-739-861A-1
3	162.5	29.9	371	10	US-09-795-583-1
4	161.5	29.7	386	10	US-09-739-861A-5
5	161.5	29.7	386	10	US-09-795-583-5
6	93	17.1	1026	10	US-09-379-931-7
7	92	16.9	596	12	US-10-006-867-100
8	92	16.9	596	12	US-10-052-586-310
9	80.5	16.3	4440	12	US-10-052-586-525
10	86	15.8	344	10	US-09-770-621-2
11	83	15.3	1477	9	US-10-092-880-4
12	80.5	14.8	619	10	US-09-800-065-2
13	76	14.0	480	10	US-09-823-936-16
14	75.5	13.9	1435	12	US-10-052-586-581
15	73.5	13.5	688	10	US-09-864-761-36047
16	73	13.4	327	10	US-09-850-351A-23
17	73	13.4	327	10	US-09-850-351A-29
18	72.5	13.4	331	10	US-09-864-761-36048
19	72	13.3	1599	9	US-10-092-880-9

20	72	13.3	1600	9	US-10-092-880-10
21	70.5	13.0	1422	10	US-09-735-933-1
22	70.5	13.0	4679	10	US-09-804-898-2
23	70	12.9	218	9	US-09-916-494A-34
24	70	12.9	234	9	US-10-075-872-3
25	69.5	12.8	263	10	US-09-347-064-10
26	69.5	12.8	267	10	US-09-347-064-4
27	69.5	12.8	605	9	US-09-741-233A-2
28	68.5	12.6	151	10	US-09-789-561-170
29	68.5	12.6	473	10	US-09-910-059-131
30	68.5	12.6	957	10	US-09-922-217-1065
31	68.5	12.6	957	10	US-09-833-263-1065
32	68	12.5	238	10	US-09-976-787-29
33	68	12.5	238	10	US-09-865-198-28
34	68	12.5	240	10	US-09-976-787-28
35	68	12.5	240	10	US-09-865-198-27
36	68	12.5	472	10	US-09-815-242-13542
37	68	12.5	479	10	US-09-815-242-13307
38	68	12.5	525	10	US-09-804-551B-36
39	67.5	12.4	234	9	US-10-075-872-4
40	67.5	12.4	450	10	US-09-796-848A-37
41	67.5	12.4	450	10	US-09-796-848A-39
42	67.5	12.4	450	10	US-09-796-848A-41
43	67.5	12.4	450	10	US-09-796-848A-43
44	67.5	12.4	450	10	US-09-796-848A-45
45	67.5	12.4	450	10	US-09-796-848A-47

ALIGNMENTS

RESULT 1

US-09-981-900B-5

; Sequence 5, Application US/09981900B

; Patent No. US20020138878A1

; GENERAL INFORMATION:

; APPLICANT: Sticklen, Masomeh B

; APPLICANT: Maqbool, Shahina B

; APPLICANT: Dale, Bruce E

; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULASE WHICH DEG

; TITLE OF INVENTION: AND CELLULOSE TO FERMENTABLE SUGARS

; FILE REFERENCE: MSU 4.1-539

; CURRENT APPLICATION NUMBER: US/09/981,900B

; CURRENT FILING DATE: 2002-03-18

; PRIOR APPLICATION NUMBER: 60/242,408

; PRIOR FILING DATE: 2000-10-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 5

; LENGTH: 562

; TYPE: PRT

; ORGANISM: Acidothermus cellulolyticus

US-09-981-900B-5

Query Match 73.5%; Score 399; DB 10; Length 562;
Best Local Similarity 69.3%; Pred. No. 3.7e-32;
Matches 70; Conservative 13; Mismatches 18; Indels 0; Caps 0;

OY	1	GVGCRTATVYVNSDMGSGFTA	VTVTNTGSRATSGWTVANSFGNGQTVNTYNTALTOSGA	60
Db	462	GARCTASYQVNSDMGNGFTVTVAVTNSGVSATKTWTVMTFGGNTITNSMNAAVTONGO	521	
OY	61	SVTATNLNYYVLPQCGSTTFGFGNGSYSGTNTAPTTLCTAS	101	
Db	522	SVTARNNSYNNVLPQCGNTTFGFGASVTGSNAAPTACAAAS	562	

RESULT 2

US-09-739-861A-1

; Sequence 1, Application US/09739861A

; Patent No. US20020076792A1

; GENERAL INFORMATION:

; APPLICANT: Jones, Brian E.

```

; APPLICANT: van der Kleij, Wilhelmus A.H.
; APPLICANT: van Solingen, Piet
; APPLICANT: Weyler, Walter
; APPLICANT: Goedegebuur, Frits
; TITLE OF INVENTION: No. US2002007692A1el Cellulase Producing Actinomycetes.
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
; FILE REFERENCE: GC540-3
; CURRENT APPLICATION NUMBER: US/09/739,861A
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: US 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: US 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" = Streptomycetes
; OTHER INFORMATION: thermobacillus
US-09-739-861A-1

Query Match          29.9%; Score 162.5; DB 10; Length 371;
Best Local Similarity 33.7%; Pred. No. 4.8e-09;
Matches 35; Conservative 19; Mismatches 41; Indels 9; Gaps 4;

Qy      4 C R A T Y V N S D M G S G C T A T V T V T N T G S R A T S G W T V A K S F C G N Q T V T N Y N T A L T O S G A S V T   63
Db      267 C Q V S I S H T - W P G G F T V D T T I N T G S T P V D G W E L D F T L P A G H T V T S V M N A L I S P A S G A V T   325
               | : : | : : | : | | | | | | | | : : | | : : | | : : | : | : | : | : |
Qy      64 A T N L S Y N - N V I Q P Q G S T T F G N G S Y G T N - T A P T - - - - - L T C T       99
Db      326 A R S T G S C R I A A N G G T Q S F G F O G T S S G A G F T A P A G A R L N G T S C T   369
               | : : | : : | : | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
US-09-795-583-1
; Sequence 1, Application US/09795583
; Patent No. US20020081702A1
; GENERAL INFORMATION:
; APPLICANT: Genecor International, Inc.
; TITLE OF INVENTION: No. US20020081702A1el Cellulase Producing Actinomycetes,
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/795,583
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" = Streptomycetes
; OTHER INFORMATION: thermobacillus
US-09-795-583-1

Query Match          29.9%; Score 162.5; DB 10; Length 371;
Best Local Similarity 33.7%; Pred. No. 4.8e-09;
Matches 35; Conservative 19; Mismatches 41; Indels 9; Gaps 4;

```

```

QY      4  CRATVYVNSDWSGSGTATVTVNTGSRATSGWTVAWSFGGNGOTVTNYWNTALTQSGASVT 63
      .  I : : : I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db     262  COVSYSTHT-WPGGFTVDTTINTGCTPVDGHELDFTLPAGHTVTSVMNALISPASGAVT 325

QY      64  ATNLSYN-NVIQPGOSTTFGNGSYSGTN-TAPT-----LTCT 99
      .  I : : : I : : : I I I I I I I I I I I I I I I I I I I I I I I I I
Db     326  ARSTGSNGRIANGCTQSGFGCTSSGAGFTAPAGARLNGTSTCT 369

RESULT 4
US-09-739-861A-5
: Sequence 5, Application US/09739861A
: Patent No. US20020076792A1
: GENERAL INFORMATION:
: APPLICANT: Jones, Brian E.
: APPLICANT: van der Kleij, Wilhelmus A.H.
: APPLICANT: van Solingen, Piet
: APPLICANT: Weyler, Walter
: APPLICANT: Goedegebuur, Frits
: TITLE OF INVENTION: No. US20020076792A1el Cellulase Producing Actinomycetes.
: TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same.
: FILE REFERENCE: GC540-3
: CURRENT APPLICATION NUMBER: US/09/739.861A
: CURRENT FILING DATE: 2000-12-18
: PRIOR FILING DATE: US 09/321,981
: PRIOR APPLICATION NUMBER: 1999-05-28
: PRIOR FILING DATE: 1999-05-28
: PRIOR APPLICATION NUMBER: US 09/104,308
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: US 08/974,042
: PRIOR FILING DATE: 1997-11-19
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 386
: TYPE: PRT
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Nearest "neighbor" - Streptomyces
US-09-739-861A-5

Query Match          29.7%: Score 161.5; DB 10; Length 386;
Best Local Similarity 33.7%: Pred. No. 6.2e-09;
Matches 35; Conservative 19; Mismatches 41; Indels 9; Gaps 4;

QY      4  CRATVYVNSDWSGSGTATVTVNTGSRATSGWTVAWSFGGNGOTVTNYWNTALTQSGASVT 63
      .  I : : : I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db     262  COVSYSTHT-WPGGFTVDTTINTGCTPVDGHELDFTLPAGHTVTSVMNALISPASGAVT 340

QY      64  ATNLSYN-NVIQPGOSTTFGNGSYSGTN-TAPT-----LTCT 99
      .  I : : : I : : : I I I I I I I I I I I I I I I I I I I I I I I I I
Db     341  ARSTGSNGRIANGCTQSGFGCTSSGTFGNAPAGRLNGTSTCT 384

RESULT 5
US-09-795-583-5
: Sequence 5, Application US/09795583
: Patent No. US20020081702A1
: GENERAL INFORMATION:
: APPLICANT: Genencor
: TITLE OF INVENTION: No. US20020081702A1el Cellulase Producing Actinomycetes.
: TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same.
: FILE REFERENCE: GC540-2
: CURRENT APPLICATION NUMBER: US/09/795.583
: CURRENT FILING DATE: 2001-02-27
: PRIOR FILING DATE: 2001-02-27
: PRIOR APPLICATION NUMBER: 09/321,981
: PRIOR FILING DATE: 1999-05-28
: PRIOR APPLICATION NUMBER: PCT/US99/11971
: PRIOR FILING DATE: 1999-05-28
: PRIOR APPLICATION NUMBER: 09/104,308
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 08/974,042
: PRIOR FILING DATE: 1997-11-19

```

```

QY      4 CRATVYVNSDMSGSGTATVTNTGSRATSGWTVAWSFGGNQTVTNYNWTALTQSCASVT 63
       I : : : I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db     267 COVSYSTHT-WPGGFTVDTTINTGSTPVDGWELDFTLTPAGHTVTSAMNALISPASGAVT 325

QY      64 ATNL SYN-NVIOPGOSTTFGFNGSYSGTN-TAPT-----LTCT 99
       I : : : I : : : I I I I I I I I I I I I I I I I I I I I I I I I I I
Db     326 ARSTGSNGRIANGTGOSFGFOGTSSGAGFTAPAGARLNGTSTCT 369


RESULT 4
US-09-739-861A-5
; Sequence 5, Application US/09739861A
; Patent No. US20020076792A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Brian E.
; APPLICANT: van der Kleij, Wilhelmus A.H.
; APPLICANT: van Solingen, Piet
; APPLICANT: Weyler, Walter
; APPLICANT: Goedegebuur, Frits
; TITLE OF INVENTION: No. US20020076792A1el Cellulase Producing Actinomycetes.
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same.
; FILE REFERENCE: GC540-3
; CURRENT APPLICATION NUMBER: US/09/739.861A
; CURRENT FILING DATE: 2000-12-18
; PRIOR FILING DATE: US 09/321,981
; PRIOR APPLICATION NUMBER: 1999-05-28
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: US 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: US 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" - Streptomyces
US-09-739-861A-5

Query Match          29.7%: Score 161.5; DB 10; Length 386;
Best Local Similarity 33.7%; Pred. No. 6.2e-09;
Matches 35; Conservative 19; Mismatches 41; Indels 9; Gaps 4;

QY      4 CRATVYVNSDMSGSGTATVTNTGSRATSGWTVAWSFGGNQTVTNYNWTALTQSCASVT 63
       I : : : I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db     262 COVSYSTHT-WPGGFTVDTTINTGSTPVDGWELDFTLTPAGHTVTSAMNALISPASGAVT 340

QY      64 ATNL SYN-NVIOPGOSTTFGFNGSYSGTN-TAPT-----LTCT 99
       I : : : I : : : I I I I I I I I I I I I I I I I I I I I I I I I I I
Db     341 ARSTGSNGRIANGTGOSFGFOGTSSGTFGNAPAGRLNGTSTCT 384


RESULT 5
US-09-795-583-5
; Sequence 5, Application US/09795583
; Patent No. US20020081702A1
; GENERAL INFORMATION:
; APPLICANT: Genencor
; APPLICANT: International, Inc.
; TITLE OF INVENTION: No. US20020081702A1el Cellulase Producing Actinomycetes, Same
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same,
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/795.583
; CURRENT FILING DATE: 2001-02-27
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19

```


;
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100684
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102570
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105881
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/106030
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/106464
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 60/106856
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: 60/108807
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/112419
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/112422
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/112853
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113011
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/112854
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113300
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113408
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113430
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621

;
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114223
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 60/115614
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/116527
; PRIOR FILING DATE: 1999-01-20
; PRIOR APPLICATION NUMBER: 60/116843
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/119285
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119287
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119525
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119549
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/120014
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/129122
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/129674
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/138387
; PRIOR FILING DATE: 1999-06-09
; PRIOR APPLICATION NUMBER: 60/144791
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/169495
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/175481
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 60/191007
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/199397
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/380139
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 09/311832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/380137
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/380138
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/380142

Query Match 16.9%; Score 92; DB 12; Length 596;

Best Local Similarity 35.0%; Pred. No. 0.071;

Matches 35; Conservative 12; Mismatches 39; Indels 14; Gaps 5;

Qy 2 VGCRTYVNVSDMGSGFTATVTVTNGSRATSGWTVMVSGGNOTVTNYWNTAL.TOSGAS, 61

Db 218 VSSRATATNSESSTTSSGASTATNSESRTTS-----NGAGTATNS-ESSTTSSGAS 268
| | | | | : | | | | | : | | | | |
| | | | | : | | | | | : | | | | |

Qy 62 VTATNLSYNNVIOPGOSTTGFENGSGTNTAPLTLCCTAS 101
||||| | : | | | | | : | | | | |

Db 269 -TATN-SDSSTVSSGASTA---TNSSESTTSSGASTATNS 303

RESULT 8

US-10-052-586-310

; Sequence 310, Application US/10052586

; Patent No. US20020127584A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Collin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052,586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05

[illegible]

1 PRIOR APPLICATION NUMBER: 60/081195
2 PRIOR FILING DATE: 1998-04-09
3 PRIOR APPLICATION NUMBER: 60/081838
4 PRIOR FILING DATE: 1998-04-15
5 PRIOR APPLICATION NUMBER: 60/082568
6 PRIOR FILING DATE: 1998-04-21
7 PRIOR APPLICATION NUMBER: 60/082569
8 PRIOR FILING DATE: 1998-04-21
9 PRIOR APPLICATION NUMBER: 60/082704
10 PRIOR FILING DATE: 1998-04-22
11 PRIOR APPLICATION NUMBER: 60/082797
12 PRIOR FILING DATE: 1998-04-22
13 PRIOR APPLICATION NUMBER: 60/083322
14 PRIOR FILING DATE: 1998-04-28
15 PRIOR APPLICATION NUMBER: 60/083495
16 PRIOR FILING DATE: 1998-04-29
17 PRIOR APPLICATION NUMBER: 60/083496
18 PRIOR FILING DATE: 1998-04-29
19 PRIOR APPLICATION NUMBER: 60/083499
20 PRIOR FILING DATE: 1998-04-29
21 PRIOR APPLICATION NUMBER: 60/083559
22 PRIOR FILING DATE: 1998-04-29
23 PRIOR APPLICATION NUMBER: 60/084366
24 PRIOR FILING DATE: 1998-05-05
25 PRIOR APPLICATION NUMBER: 60/084414
26 PRIOR FILING DATE: 1998-05-06
27 PRIOR APPLICATION NUMBER: 60/084639
28 PRIOR FILING DATE: 1998-05-07
29 PRIOR APPLICATION NUMBER: 60/084640
30 PRIOR FILING DATE: 1998-05-07
31 PRIOR APPLICATION NUMBER: 60/084643
32 PRIOR FILING DATE: 1998-05-07
33 PRIOR APPLICATION NUMBER: 60/085573
34 PRIOR FILING DATE: 1998-05-15
35 PRIOR APPLICATION NUMBER: 60/085579
36 PRIOR FILING DATE: 1998-05-15
37 PRIOR APPLICATION NUMBER: 60/085580
38 PRIOR FILING DATE: 1998-05-15
39 PRIOR APPLICATION NUMBER: 60/085582
40 PRIOR FILING DATE: 1998-05-15
41 PRIOR APPLICATION NUMBER: 60/085700
42 PRIOR FILING DATE: 1998-05-15
43 PRIOR APPLICATION NUMBER: 60/086023
44 PRIOR FILING DATE: 1998-05-18
45 PRIOR APPLICATION NUMBER: 60/086392
46 PRIOR FILING DATE: 1998-05-22
47 PRIOR APPLICATION NUMBER: 60/086486
48 PRIOR FILING DATE: 1998-05-22
49 PRIOR APPLICATION NUMBER: 60/087098
50 PRIOR FILING DATE: 1998-05-28
51 PRIOR APPLICATION NUMBER: 60/087208
52 PRIOR FILING DATE: 1998-05-28
53 PRIOR APPLICATION NUMBER: 60/087609
54 PRIOR FILING DATE: 1998-06-02
55 PRIOR APPLICATION NUMBER: 60/087759
56 PRIOR FILING DATE: 1998-06-02
57 PRIOR APPLICATION NUMBER: 60/087827
58 PRIOR FILING DATE: 1998-06-03
59 PRIOR APPLICATION NUMBER: 60/088025
60 PRIOR FILING DATE: 1998-06-04
61 PRIOR APPLICATION NUMBER: 60/088028
62 PRIOR FILING DATE: 1998-06-04
63 PRIOR APPLICATION NUMBER: 60/088029
64 PRIOR FILING DATE: 1998-06-04
65 PRIOR APPLICATION NUMBER: 60/088033
66 PRIOR FILING DATE: 1998-06-04
67 PRIOR APPLICATION NUMBER: 60/088167
68 PRIOR FILING DATE: 1998-06-05
69 PRIOR APPLICATION NUMBER: 60/088202
70 PRIOR FILING DATE: 1998-06-05
71 PRIOR APPLICATION NUMBER: 60/088212
72 PRIOR FILING DATE: 1998-06-05
73 PRIOR APPLICATION NUMBER: 60/088217

1 PRIOR FILING DATE: 1998-06-05
2 PRIOR APPLICATION NUMBER: 60/088326
3 PRIOR FILING DATE: 1998-06-04
4 PRIOR APPLICATION NUMBER: 60/088655
5 PRIOR FILING DATE: 1998-06-09
6 PRIOR APPLICATION NUMBER: 60/088722
7 PRIOR FILING DATE: 1998-06-10
8 PRIOR APPLICATION NUMBER: 60/088738
9 PRIOR FILING DATE: 1998-06-10
10 PRIOR APPLICATION NUMBER: 60/088740
11 PRIOR FILING DATE: 1998-06-10
12 PRIOR APPLICATION NUMBER: 60/088811
13 PRIOR FILING DATE: 1998-06-10
14 PRIOR APPLICATION NUMBER: 60/088824
15 PRIOR FILING DATE: 1998-06-10
16 PRIOR APPLICATION NUMBER: 60/088825
17 PRIOR FILING DATE: 1998-06-10
18 PRIOR APPLICATION NUMBER: 60/088826
19 PRIOR FILING DATE: 1998-06-10
20 PRIOR APPLICATION NUMBER: 60/088861
21 PRIOR FILING DATE: 1998-06-11
22 PRIOR APPLICATION NUMBER: 60/088863
23 PRIOR FILING DATE: 1998-06-11
24 PRIOR APPLICATION NUMBER: 60/088876
25 PRIOR FILING DATE: 1998-06-11
26 PRIOR APPLICATION NUMBER: 60/089090
27 PRIOR FILING DATE: 1998-06-12
28 PRIOR APPLICATION NUMBER: 60/089105
29 PRIOR FILING DATE: 1998-06-12
30 PRIOR APPLICATION NUMBER: 60/089512
31 PRIOR FILING DATE: 1998-06-16
32 PRIOR APPLICATION NUMBER: 60/089514
33 PRIOR FILING DATE: 1998-06-16
34 PRIOR APPLICATION NUMBER: 60/089538
35 PRIOR FILING DATE: 1998-06-17
36 PRIOR APPLICATION NUMBER: 60/089598
37 PRIOR FILING DATE: 1998-06-17
38 PRIOR APPLICATION NUMBER: 60/089653
39 PRIOR FILING DATE: 1998-06-17
40 PRIOR APPLICATION NUMBER: 60/089908

Query Match 16.3% Score 88.5; DB 12; Length 4440;
Best Local Similarity 27.4%; Pred. No. 1.4;
Matches 31; Conservative 9; Mismatches 56; Indels 17; Gaps 2;

Qy 3 GCRATYVYVNSDWGSGFTATVTNTGSRATSGWTVAHSFG-----GNOTV 47
Db 4094 GAAATCTTCACTGAGACTTATTTTGGCTTT--TCACATGCTCCTCCAAATGATTTGTTAA 4151
Qy 48 TNYWNTALTQSGASVTATNLNINNVIOPGQSTTTTCFNGSYSGTNTAPTLLCTA 100
Db 4152 ATTTATTTTAGGTATTTTAGGGGAATGATTTTCTAAAGTTTGTATTTCTTA 4204

RESULT 10
US-09-770-621-2
Sequence 2, Application US/09770621
Patent No. US20010024815A1
GENERAL INFORMATION:
APPLICANT: M ntyl , Arja
APPLICANT: Vehmaanper , Jari
APPLICANT: Fagerstr m, Richard
APPLICANT: Lantto, Raija
APPLICANT: Paloheimo, Marja
APPLICANT: Suominen, Pirkko
APPLICANT: Lahtinen, Tarja
TITLE OF INVENTION: Production and Secretion of Proteins of
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 New York Ave., N.W. Suite 600
CITY: Washington
STATE: D.C.


```
; ORGANISM: P. Yoell
; US-09-823-936-16

Query Match      14.0% Score 76; DB 10; Length 480;
Best Local Similarity 30.0% Pred. No. 2.1; Mismatches 36; Gaps 6;
Matches 36; Conservative 5;

Oy 3 GCRATYVNSDMGSGFTA-----TVTVTNTGSRATSGMTVAWSFG----GNQTVTNW 51
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 361 GCCATAAGGAGCGCATATAATTTAATACTAGTAATTGATCAGAGGAACCGGATCTT-- 418

Oy 52 NTALTQSGASVTATNLNYSNNVIOGQSTTTCFNGSYS-----GTNTAPTLT-----CTAS 101
   ||| ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 419 ATAATCGGAGCTCT-----TTCGAGAGCTTCTCTGCTTTGGTTTGGACCTCT 465

RESULT 14
US-10-052-586-581
; Sequence 581, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11

; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
```


; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
US-09-864-761-36047

Query Match 13.5%; Score 73.5; DB 10; Length 688;
Best Local Similarity 32.0%; Pred. No. 5.5;
Matches 32; Conservative 12; Mismatches 43; Indels 13; Gaps 4;
QY 3 GCRATYVNVSDMGSGFTATVTVTGSRATSGWTVAWSFGNGQVTNYYNWTALTQSGASV 62
Db 460 GSEATTVSTTS-----SETTTASTTGSEMT---TVFTTVSETTVSTIGSEATTSSAAGS 511
QY 63 TATNLSYNNVLOPGOSTFGNGSYSGT-NTAPTLTCTAS 101
Db 512 EATTTS-----TEGSETTTASTAGSETTTASTAGSETTTAS 547

Search completed: November 13, 2002, 12:08:37
Job time : 2.95905 secs

